Univ. of
Toronto
Library

BINDING LIST DEC 1 1924

Digitized by the Internet Archive
in 2007 with funding from
Microsoft Corporation

http://www.archive.org/details/no2t17bulletin00univuoft

Vol. XVII. October 27, 1919 No. 9

BUREAU OF EDUCATIONAL RESEARCH—BULLETIN No. 2

Bureau of Educational Research

COLLEGE OF EDUCATION
First Annual Report
Announcement, 1919-20

193339
8.1.25

Price, 25 cents

PUBLISHED BY THE UNIVERSITY OF ILLINOIS
URBANA

LB
1028
A1145
no.2-17

TABLE OF CONTENTS

Chapter		Page
I.	Organization	7
II.	Projects undertaken	13
III.	Proposed projects	31
IV.	The work of the library division	39
V.	Measuring the results of teaching for the purpose of improving instruction	43
VI.	Standardized tests for the elementary school	47
VII.	Standardized tests for high school subjects	64
VIII.	Mental tests	67
	Directions for ordering tests and order blank	71

FOREWORD

This bulletin is at once a yearbook embodying the work of the Bureau of Educational Research last year and an announcement indicating the projects which the Bureau contemplates and the test business which it is prepared to do. Acknowledgements are made to Mr. H. T. McKinney in connection with the promotion project, to Rev. John A. O'Brien in connection with the reading project, to Miss Velda Bamesberger in connection with the project on memory work in the grades, and to Dr. C. E. Holley in connection with the try-out of group tests of intelligence. These projects are briefly described in Chapter II. In general our indebtedness is very great to the school people of the state for the cooperation they have accorded to the Bureau in the first year of its work.

November, 1919 B. R. Buckingham

Director

CHAPTER I. ORGANIZATION.

Personnel and equipment. The Bureau of Educational Research of the University of Illinois was created with the appointment of a director in July 1918. No provision had been made at that time in the way of rooms, equipment, or assistance. The director occupied space in a room with three others, and his desk was borrowed. So also was the stenographer who attended to his mail.

Nor were rooms, materials, or personal service easily obtainable. The war was on, and all available space in university buildings was being utilized by the aviators and afterwards by the Students Army Training Corps. As to equipment and service, while ample funds had been provided for them, they were almost impossible to obtain. Experienced stenographers and clerks were not to be had, and it was a long time before even those without experience could be secured and their training begun. Deliveries of equipment were very uncertain, were generally long delayed, and sometimes failed altogether. Thus to the usual difficulties of organizing such a venture as the Bureau represents were added far greater difficulties due to the time when the organization took place.

Nevertheless, now after a little more than a year the Bureau is occupying five rooms in which eleven full-time and three half-time employees are working. The Bureau personnel includes a director and assistant director, two faculty members of the rank of assistant, two library assistants, three stenographers, and two clerks. The three half-time employees are students. One of them is a stenographer and the other two are shipping clerks.

A great deal of time and energy was used during the first year of the existence of the Bureau in building up this organization and in the purchase and installation of the $4000 worth of equipment now in use. Yet a great deal more was done than simply to organize. Perhaps the surest indication of this is the amount of mail matter received by and sent from the Bureau. Besides great quantities of other than first class matter, there were during the year ended June 30, 1919, 2918 incoming letters and 3674 outgoing letters. This was for a period of about ten months.

Reference Library. The collection and proper handling of the sort of printed material which the Bureau needed to have at its disposal, constituted a large part of the work during this, the first year of organization. At present we have about 400 courses of study, 300 city and state educational reports, the school laws of all the states which publish such compilations, 138 volumes of surveys, and about 3000 bulletins and pamphlets of various kinds. Fifty-nine periodicals are regularly received by the Bureau.

Educational Tests and Measurements. Many activities of the Bureau have been in connection with standardized test material. In this field five lines of work may be noted: (1) publication and purchase of test materials; (2) the preparation of accessory material for the administration and scoring of the tests; (3) distribution of these materials to the schools; (4) the derivation of new tests; and (5) correspondence relative to the use of test materials by teachers and superintendents.

(1) Publication and Purchase of Tests. The Bureau publishes far the greater part of the test material which it distributes. No other similar bureau does this to so great a degree. The Bureau is the sole publisher of the following tests including in each case directions, score sheets, and other necessary accessories:

1. Buckingham's Extension of the Ayres Spelling Scale.
2. Buckingham's Scale for Problems in Arithmetic, Divisions 1, 2, and 3.
3. Charters' Diagnostic Language Tests, Pronouns, Verbs A, Verbs B, and Miscellaneous.
4. Charters' Diagnostic Language and Grammar Tests, Pronouns, Verbs A, and Miscellaneous.
5. Cleveland Survey Arithmetic Tests.
6. Harlan's Test for Information in American History.
7. Holley's Picture Completion Test for Primary Grades.
8. Holley's Sentence Vocabulary Test.
9. Monroe's Diagnostic Tests in Arithmetic, Parts I, II, III, and IV.
10. Monroe's Standardized Reasoning Tests in Arithmetic, Tests I, II, and III, forms 1 and 2.
11. Monroe's Standardized Silent Reading Tests, I, II, and III.
12. Monroe's Timed Sentence Spelling Tests I, II, and III.
13. Willing's Scale for Measuring Written Composition.

Permission has also been given to the Bureau to print and distribute the following, together with accessories:

14. Gray's (C. T.) Standard Score Card for Measuring Handwriting, Forms I, II, III, and IV.
15. Gray's (W. S.) Oral Reading Test.
16. Gray's (W. S.) Silent Reading Tests.
17. Starch's Punctuation Scale.
18. Thorndike's Scale Alpha 2: Understanding of Sentences.
19. Thorndike's Visual Vocabulary Scales.

The Bureau likewise publishes score sheets, directions, standards, and answer sheets for a number of other tests and scales which it purchases from other publishers. Among those for which some or all of these accessories are published by the Bureau are the following:

1. Ayres' Handwriting Scale.
2. Ayres' Spelling Scale.
3. Freeman's Chart for Diagnosing Faults in Handwriting.
4. Hahn-Lackey Geography Scale.
5. Henmon's Latin Tests.
6. Hotz' First Year Algebra Scale.
7. Minnick's Geometry Tests.
8. Nassau County Supplement to the Hillegas Composition Scale.
9. Sackett's Scale in Ancient History.
10. Sackett's Scale in United States History.
11. Starch's Physics Tests.
12. Thorndike's Handwriting Scale.
13. Trabue's Language Scales.
14. Woody's Arithmetic Scales.

Besides the tests listed above, the Bureau also purchases from other publishers certain tests which are completely equipped with the accessory material needed in using them. The most conspicuous ones are the various Courtis tests—Arithmetic, Series B, Supervisory Test in Geography, and the Silent Reading Test, No. 2.

(2) Preparation of accessory materials. As may be inferred from the above statement, the authors of a considerable number of standardized tests have not provided directions for administering and scoring them or score sheets on which results may be systematically recorded. The Bureau has been engaged during the past year in making good these deficiences, until it is now able to announce that every test which it offers for sale will be furnished with all the necessary instructions and forms for using it.

(3) Distribution of tests. Test material has been sold in every section of the United States. Naturally, orders came from Illinois in greater numbers than from any other state. The total sales up to the 30th of June, 1919 amounted to about 77,000 copies of tests of an approximate value of $900.00. Now, however, it appears that this was only a beginning. At the present writing (October 15th, 1919) the Bureau has already received orders for three or four times as many tests as were sold during the whole of last year.

Since the great volume of orders clearly indicates that a distinct service can be rendered to superintendents and teachers by making easily accessible standardized tests accompanied by the necessary accessories, the Bureau is undertaking to render such service. Since the project is assuming considerable dimensions it is necessary to put it on a business basis. The retail prices of test materials includes a sufficient advance over the printing or purchase cost to defray the expense for stenographic and clerical service required for handling the materials. The resources of the Bureau are such that it is able to print or purchase tests in large quantity thus securing minimum prices and maintaining a stock of sufficient size so that even orders for large numbers of tests can be filled at once.

(4) Derivation of tests. No small attention has been given to evaluating new test material. Some of this material has been printed; while some of it still requires much more cooperative effort before it can merit publication. Returns were received from over two hundred superintendents and principals involving more than 50,000 pupils. A more detailed account of these projects will be given in the next chapter.

(5) Correspondence concerning tests. A considerable correspondence has always been carried on concerning the tests. Sometimes orders are ambiguous and we have to write for fuller information. Sometimes orders cannot be filled completely and letters explaining the situation have to be written. Advice is frequently asked as to what test to use, as to the meaning of certain test-questions, what answers to regard as correct, what score should be given under certain circumstances, what the meaning of the results is, and what can be done to remedy defects. Someone in the Bureau always sends a reply in which an attempt is made to furnish the desired information. Some of the inquiries have been very suggestive and have resulted in the preparation of useful directions and explanations to accompany the tests.

Organization of Bureau into divisions. Thus as the organization of the Bureau progressed during the year 1918-19, it did so with particular reference to two purposes. The first was to obtain and properly handle as rich a collection of research material as possible. The second was to distribute the best test material and to assist school people in its use and interpretation. It is natural, therefore, that as the second year of the life of the Bureau begins, these two major lines of work should be the bases for the first organizations of the Bureau into divisions. We now have the Division of Educational Tests and Measurements at the head of which is Dr. Walter S. Monroe, the assistant director of the Bureau, and the Library Division which consists of two library assistants who catalog and index the printed material received by the Bureau. They also prepare bibliographies and assesmble evidence on problems submitted to the Bureau by the school people of the state and by university students and faculty members.

It is probable that this divisional organization of the Bureau will be carried out more completely as the Bureau develops and meets more fully the educational needs of the state. What these educational needs are has been in part revealed by the school people themselves. In November 1918 and again in April 1919 a circular letter and questionnaire were sent to all superintendents and high school principals of the state and to principals of elementary schools in a few of the larger cities. Four hundred forty-eight replies were received. These have afforded and will continue to afford a sure basis of action. Almost all the writers said that they were interested in the experimental study of education, that they wished to unite with the school men and women of the state in such study, and that the Bureau could help them in ways which they specifically indicated. They also listed the topics which they would like to study. Since far the greater part of these topics concerned some phase of educational tests, it was natural that first consideration should be given to this expression of interest. Hence the organization of the Division of Educational Tests and Measurements under Dr. Monroe. The Library Division constituted a service line equally available for all types of work. What the Bureau has done and hopes to do through this division will be set forth in another place.

Intelligence testing. The responses of the school men and women of the state clearly indicated the need for other types of work than those relating to educational tests and measurements. As these other types of work develop, it is to be expected that other divisions of the Bureau will be created. For example, the demand for intelligence tests in the schools is becoming more and more insistent. With the devising of simple intelligence tests which may be given simultaneously to large numbers of children—tests which at the same time seem to be more reliable than any group tests we have hitherto used—their availability for school use becomes apparent. As long as we were obliged to test children individually and the work required a high degree of expertness, the idea of basing action with reference to school children upon their mentality, although it was recognized as important, could not be carried out in practice. Now, however, it is not too much to expect that wide-awake principals and superintendents will require that all children in their schools be mentally tested upon entrance to school and periodically thereafter. At present this demand is affecting the Bureau only in a small way. We are offering this year four group intelligence scales for sale. So far as we know, this is the first time a bureau has announced that it will distribute such tests. These intelligence tests are described in another part of this bulletin. A description is also given of other group intelligence tests which may be purchased elsewhere.

It is probable, however, that the time will shortly come when it will be necessary to set up a division of intelligence tests either within the Bureau or in the Department of Educational Psychology or by some co-operative arrangement between that Department and the Bureau.

School accounting. Similarly the demand for a better system of school accounting is likely to develop this type of work beyond the ability of the Bureau to handle it without a definite organization for that purpose. Accordingly, an accounting division may not inappropriately be created.

School housing. Again, requests for advice as to schoolhouse construction, ventilation, and sanitation, became particularly insistent during the spring and summer of last year. This demand does not so often come from professional educators as from Boards of Education and building committees. Several hundred school houses are built each year in the state of Illinois. For the most part, the construction is under the direction of architects who have no special knowledge of the problems of school housing. Unless some other agency of the state takes care of the need arising from this situation, the Bureau can render a very real service to the cause of education if it secures the services, at least a part of the time, of an architectural engineer. Thus may be created another division of the Bureau.

Whether any or all of these divisions—or other divisions not mentioned—will be organized will depend wholly upon the amount and character of the needs of the schools. These needs will be expressed in the form of requests for service. The Bureau will not be justified in developing a specific functional division unless the demand is strong for the service which such a type of organization alone can perform.

CHAPTER II. PROJECTS UNDERTAKEN

Organization work last year did not prevent the undertaking of several research projects. While most of these will be made on the basis of subsequent reports they will be briefly described in this chapter.

PROMOTION STUDIES.

Promotion as an administrative problem. There is evidence to show that the present method of promoting children is unsatisfactory. In particular it is evident that it disregards to an astonishing degree differences in ability among children. With the movement to measure the product of instruction we have come into possession of a vast body of data on this question of individual differences. Years ago we knew that these differences existed, but we did not know how great they were. We have now become familiar with the notion that in a given subject twenty or thirty percent of the children in a grade will show ability equal to or greater than that of typical children in the next higher grade, or that ten percent of them will surpass typical children graded two years above them. Yet it is difficult to find a school system which uses such information to secure a better classification of children.

There is also evidence that children do not profit by repeating the work of a grade. Professor H. L. Smith of Indiana University in his "Survey of a City School System" gives figures to show that children spending a second year in a grade do not obtain higher marks than they did during the first year. Our own investigations show that although teachers confidently expect children to do better work when repeating a grade,—and indeed keep them back in order that they may,—the majority of them fail to do so. While, therefore, it is true that in addition to children of higher-grade ability there are also children of lower-grade ability in every class, this is not in itself a sufficient reason for holding them back or for demoting them. It may be a reason for placing them in some kind of a special class better suited to their needs than the regular class. But is seems likely that the intellectual loss due to a repitition and the moral loss due to a sense of failure are the only inevitable results of a policy of holding children back or of demoting them.

Trial promotion. There are certain means at hand, even in schools quite unaffected by the results of experimental methods, for preventing this intellectual and moral loss. No school, for example, needs to base its decision as to the promotion of a child merely upon his record in the grade prior to that to which promotion is contemplated. A child upon completing the work of the sixth grade becomes a candidate for advancement to the seventh grade. If he secures the promotion, he will often have a new teacher and a different room. He will certainly be brought in contact with a richer curriculum. It may be a question of entrance into a Junior High School. How then shall we decide whether this sixth grade child shall be promoted to the seventh grade? Shall we do so on the basis of his sixth grade work or on the basis of a trial in the seventh grade? Suppose the child in question has not done the sixth grade work very well—perhaps has not "passed" the final examination in the sixth grade. Are we to require him to do the sixth grade work over again with the chances about even than he will do it no better than he did it before? Or shall we give him a trial in the new grade with its new teacher and more interesting course of study? Shall we consult the needs of the child or the grade standards of our schools? Shall we decide the question of a child's advancement on the basis of what he has learned or on the basis of what he needs to learn?

Two promotion projects undertaken. In deciding to make a study of promotion as an administrative question two points of view are adopted. The *first* has to do with individual differences and the *second* with the advancement of all pupils to the next higher grade and the special treatment of doubtful pupils during a probationary period. The first of these points of view was adopted in an investigation carried out at Danville, while the second was exemplified in a study made at Decatur and Springfield. The deails of these investigations were largely in the hands of Mr. H. T. McKinney, a graduate student in education.

(1) Individual difference method. Superintendent G. P. Randle of Danville consented to have given toward the close of the first semester of 1918-19 a series of tests to the children in grades three to eight of all the elementary schools in the city. This series consisted of three types.

The first was designed to give data concerning the child's proficiency in the curriculum as taught at Danville. Each child in the lower fifth grade for example, was given the arithmetic test pertaining to his grade. If he passed, he was allowed to take one pertaining to the upper fifth grade, and so on as far as his success permitted him to go. Thus, in terms of the curriculum, a child would show title to promotion in arithmetic to the grade next higher than the last one whose test he passed successfully.

The second type of testing was made with standardized material in the major subjects. According to their performing on these tests, children were rated as possessing the ability of the grade whose standard they most nearly approached.

The third type of testing consisted of half of the Pressey Group Point Scale. This was designed to measure the intelligence of the pupils.

Promotion results. With data from these sources in hand, with the teachers' estimates of the abilities of the pupils, and with the knowledge of their chronological ages, it was felt that an intelligent selection of pupils for additional promotion could be made. Unfortunately, owing to a number of delays and to insufficient clerical help, the final results were not available till long after the mid-year promotion. Accordingly it did not seem advisable to advance a very large number of pupils an additional grade. All of those who were candidates for such additional advancement had already been promoted one grade by the regular process.

Nevertheless, the testing at Danville was not without results in the form of additional promotions. For example, nine children were promoted from the upper third grade to the lower fourth grade in March after having secured their regular promotion at the beginning of February. In June all of these nine were again promoted thus offering conclusive evidence that they were sustaining themselves after their extra promotion. Miss Nelle Haley, the primary supervisor at Danville, reports that these nine children "now stand in the upper fourth of the 4A class." Only two children were promoted from the lower fourth grade to the upper fourth grade but both of them made good again in June and are now among the best in their class.

The largest number of promotions was made from the upper fourth grade to the lower fifth grade, exclusive of those in the Lincoln School. There were thirty of these, twenty-eight of whom stood high on the promotion lists in June. At the Lincoln School, due to the crowded conditions, the ten children selcted for promotion from the upper fourth grade to the lower fifth grade could not be advanced in March. They were told, however, that a double promotion would be given them in June if their work kept up to the March standard. In every case the double promotion was made and these ten children are leading 5A class at the present time.

We wish to repeat the investigation of the possibilities of additional promotion based on individual differences. Whether this is done again at Danville or elsewhere we shall be able greatly to improve our method. More group intelligence tests are now available than were available last year. The curriculum testing should be abandoned. It serves no useful purpose except to exhibit the fact that some children, in spite of the new subject matter introduced in higher grades, are able to pass a satisfactory examination on it. The writer believes that a good group intelligence test will alone afford a reasonable basis for the placing of capable pupils in grades higher than those in which they are now classified. Not many mistakes will be made, and those which are made will not be serious in their consequences.

(2) Probationary promotion method. The promotion project as it was carried out with the co-operation of Superintendent Allen at Springfield and Superintendent Engleman at Decatur contemplated, as has been said before, a probationary period during which children who would not normally secure promotion were given an opportunity to sustain themselves in the next higher grade. This plan was first put into effect at the close of the first semester of 1918-19. Teachers were requested to make out the usual lists of promoted and non-promoted pupils. After these lists were in the hands of the superintendent then, and not until then, teachers were informed that all children were to be placed in the next higher grade. The probationary period for pupils not originally recommended for promotion lasted six weeks, during which time every effort was made on the part of the teachers to cause the probationary pupils to make good their promotions.

Right here we see a marked change of emphasis and a new result. In recent years, during which a very general effort has been made to reduce retardation and elimination, teachers who secure a high promotion rate are commended. Little account, however, is taken of what becomes of children after they have been promoted. The receiving teacher gets rid of them by one means or another; and it is not unusual for her to complain of the preparation which the children have for the work of her grade. According to the scheme as administered at Decatur and Springfield, it is the receiving teacher rather than the sending teacher who must make the effort to secure a high percentage of promoted pupils. Moreover, it is to be observed that the effort is made by the teacher with the subject matter and in the grade having to do with the present work of the child.

Results of project. Accordingly, it was not surprising that the teachers in these two cities put forth unusual effort to cause the children to maintain themselves in the grade to which they had been provisionally advanced. An individual record card was provided for each probationary pupil and on this card a record was kept by the teacher. The most interesting part of this record consisted of a notation of the devices and methods used by the teacher to secure extra effort on the part of the pupil or of his parents. One device which was quite generally used consisted of a daily report from the teacher to the parents. This report was signed by the parent and returned the next morning.

As a result of this method about seventy-five percent of the children at Decatur and Springfield, who were not recommended for promotion by the sending teacher, and who according to the usual plan would not have been promoted, were retained in the higher grade by the receiving teacher. The total promotion rate was over ninety-five percent. Mental tests were subsequently given to the five percent who failed of promotion and as might have been expected many of them were found to be of inferior mental ability. A careful analysis of these faults is being made with a view to recommending special treatment for these cases.

When the question was raised whether the same plan should be followed in June as had been followed in February, Superintendent Allen of Springfield sent a questionnaire to his teachers—a questionnaire which they were asked to make out without signing. The teachers unqualifiedly recommended certain features of the plan, made valuable criticisms of some of the other features, and in a few cases raised objections involving especially observations on the extra work required. Of the 146, however, who answered the question "Do you favor continuing the new plan?" 136 said "Yes" and 10 said "No." Some very interesting special cases of the benefit which individual children had derived from the plan were given. A similar attitude toward the plan was taken by the teachers of Decatur. Accordingly, in June it was again put into operation in both cities. At this writing the corresponding probationary period has just ended. The results, therefore, of the June promotions are not yet available. It is interesting, however, to note that at Springfield of 775 pupils listed for failure in February but subsequently promoted after the probationary period, 370, or nearly one-half, were unconditionally recommended for promotion in June and thus secured their advancement without the probationary period which they passed through in February· In other words, these 370 pupils had not only been promoted in February after being regarded as failures but had also been promoted in June—and this time without question. As to the Decatur children, Superintendent Engleman reports that *more than half* of the February probationary pupils who were advanced to a higher grade were recommended unconditionally for promotion in June.

It is evident that the value of such a method of promoting pupils cannot be determined from a single application of it. Each added semester during which the superintendents and teachers at Decatur and Springfield operate the plan will add to the value of the conclusions which can be based upon it. Perhaps all we are justified in saying at this point is that the method has been successful so far and that it offers interesting and important possibilities.

RATE OF SILENT READING

The social demand for rapid silent reading. It is easy to maintain that reading is the most important subject which a child learns while at school. It functions in all his study and it is the one product of instruction which he will surely use in adult life. The amount of reading which both the child and the adult are required to do is very great and is increasing. Of the making of many books there is no end, and the time required to read them bulks surprisingly large in the life of a civilized human being. The literary or professional man must labor unceasingly to keep abreast of the work in his field. Even a child has his "assigned readings," and in addition to these he may investigate a large assortment selected by himself.

Probably, therefore, no service which the school could perform would be more gratefully received or more instantly approved than the training of children to read rapidly without sacraficing comprehension. It is a matter of common observation that people differ widely in the rate with which they can read, and it is almost as generally recognized that those who read rapidly do not do so at a sacrifice of their understanding of what they read. There are indications also that almost anyone may be taught to read more rapidly than he is now reading.

This aspect of the problem of reading is not receiving much attention. As understood in the schools, to read well generally means to do so orally without mispronunciation and with a well modulated voice—perhaps even with the elecutionary effect. In life outside of the school room and after school days are over there is little use for this kind of reading. As we say, it has little social value. With silent reading the case is quite different. Both children and adults use constantly whatever skill they have acquired in this art. If it is true that a person's mastery of thought and theory in his chosen field depends, as indeed it does to a large extent, upon the degree to which he has covered the literature of his field, then the problem of devising a type of training in reading which will make this mastery possible at a minimum expenditure of time is urgent.

It seemed appropriate, therefore, that the Bureau should investigate the extent to which the schools can impart more efficient habits of reading, and more particularly how pupils can be taught to read more rapidly.

How the eyes move in reading. In this connection an analysis of the act of reading is important. From photographic records of eye movements we know that in sweeping from left to right along the printed line the eye pauses several times. We know too, that the movements of the eye between pauses are so rapid that reading does not take place during the movement—in other words, that we read only while the eye is at rest. Our records together with those of other investigators seem to indicate that almost the entire time of reading—certainly as much as 95 per cent of it—is consumed during the pauses which the eye makes.

Evidently, then, the fixation pauses which really are the core of the reading process must be influenced in some way if the rate of silent reading is to be accelerated. Any practicable decrease in the total of reading time must be accompanied either by a lessening of the number of pauses or by a shortening of each pause or by both these processes. Here we are clearly in the realm of habit-formation. Early in their school career children form certain eye habits which they use and continue to use in reading. There is evidence to show that children do not materially increase their rate of careful reading after they have completed the sixth grade. What we know of the formation of habits leads us to believe that there are lower order habits and higher order habits. The rapid readers whom we know have been able to replace the former by the latter. They see more at each fixation or pause in the line. Perhaps in extreme cases they see an entire line at a single pause. The amount of printed material which the fastest readers grasp at each pause is apparently well within the field of clear vision. No physiological limit seems to exist to prevent improvement in the rate of reading even on the part of the most rapid readers.

How the experiment was conducted. Superintendents and principals in nine different cities in Illinois participated in an investigation which was intended to throw light on the possibility of teaching children to read rapidly. Records were obtained from about 1600 pupils ranging from the third to the eighth grade. Each child was tested for rate of silent reading with Form I of the Courtis Silent Reading Test, No. 2. His comprehension record was also taken. On the basis of the rate records each class was divided into two groups of children whose scores were approximately the same. One of these groups was the experimental group and the other the control group. To the experimental group reading was taught according to a method which was designed to increase the rate. To the control group reading was taught in the usual manner.

The details of the special method by which the experimental group was taught will be fully reported in a future bulletin by Reverend John A. O'Brien, who was immediately in charge of the investigation. The method was outlined in general before the work began and at a meeting of representatives from all the schools concerned the details were worked out. Each teacher kept careful daily records and a dairy. Each pupil in the exprimental group kept a daily graphic record of his rate of reading.

The experimental period lasted 39 days and the teaching time was 30 minutes each day. On the 19th day of the period Form 2 of the Courtis Silent Reading Test, No. 2 was given to both the experimental and control groups. Form 3 of the same test was given to both groups on the 39th day.

Thus, besides the individual charts showing rate of silent reading in daily recitations, records for each child were also obtained in the Courtis Reading Test at the beginning, middle and end of the experimental period.

How it worked. It not infrequently happens that an experiment, after extending over a long time and involving considerable expense, yields negative results. If two methods are being compared it frequently happens that neither of them appears to be better than the other. This was emphatically not the case in this study. The children in the experimental group improved their reading rate and surpassed the children in the control group to an astonishing degree. Indeed, the improvement in rate of silent reading looked "too good to be true." In the case of one teacher who made an admirable report but whose figures showed an almost unbelievable amount of improvement for the pupils in her experimental group, some pains were taken to check up the figures. No reason, however, appeared to exist for doubting them except the fact that they were altogether at variance with our usual experience.

We, therefore, feel justified in saying that by means of a special method similar to the one used in this investigation, the rate of silent reading for the child of average ability may be greatly increased without sacrificing comprehension. How great this increase may be is yet a question, but there are certainly well established instances in which during the 39 days of practice the rate of silent reading was more than doubled. It is probable that all of this increase in rate will not be maintained. The practice period was altogether too short to ensure a lasting acquisition of the higher order habits to which we have referred, but in school practice the method could be carried out for any desired length of time and the higher order habits could be made a permanent possession. The very least that can be concluded from this investigation is that the ordinary method of teaching reading is vastly less effective in imparting rate of silent reading than a method specifically designed for that purpose. Whether in 39 days a pupil may double his rate while losing nothing in comprehension is not the vital question. The real questions are *first,* is it desirable that rapid silent reading be taught in the schools; *second,* if so, can a method be devised which will be clearly superior to the usual method of teaching reading. Our investigation enables us to answer both these questions in the affirmative.

DERIVATION OF TESTS

As a result of the questionnaire which was sent out last year to school superintendents and principals, it was possible to group together those who expressed interest in similar problems. Many indicated their fields of interest by naming a school subject, as reading or arithmetic. In fact it was not difficult to form a group for each of the principal elementary school subjects.

Of course these different groups of school people might be appealed to in any one of a number of ways. Instructional, supervisory, and administrative problems might be, and in some cases were, studied with reference to these subjects of interest. But the development of a larger and richer body of test material for certain of the subjects was thought to be particularly important. It was felt, in other words, that one function of the Bureau ought clearly to be the creation of additional standardized material.

Accordingly, a few superintendents and principals who had expressed interest in arithmetic, geography, and history were approached last fall and requested to submit to their pupils certain series of questions in these subjects. The number of schools involved at that time was not large owing to the effect of war work and of the influenza epidemic. Last April, however, when the same or similar material was again ready for distribution, a very much greater number of superintendents and principals gladly undertook the work. In this instance grammar was added to the subjects above mentioned.

Arithmetic. The material in this subject consisted of verbal problems. Some were sent out to the schools in November, many more in April. The following are some of the significant figures:

Number of superintendents and principals approached	293
Number to whom material was sent	188
Number who returned material	133
Number of problems submitted	232
Number of children from whom returns were received	31,480

This material was rated by teachers according to the directions and answer cards submitted; and all of it has now been assembled and tabulated. Those from whom material was received were furnished with a report indicating not only the scores of their children but also the scores of all the other children who answered the same questions.

At intervals during the period of three or four years prior to this testing the writer had been offering most of these problems to such superintendents and principals as were willing to cooperate with him. The new data afforded by the testing in November and April of last year seemed sufficient to warrant the formation of a scale. This scale has now been published by the Bureau of Educational Research under the title of "A Scale for Problems in Arithmetic." It consists of three divisions. The first is offered for grades three and four, the second for grades five and six, and the third for grades seven and eight. The easiest problem in the scale is: "We learn two words a day in our class. How many do we learn in 8 days?" The hardest problem is: "Mr. Wood had 250 baseballs. He sold 20% of them the first day and 40% of the remainder the second day. What per cent of his original stock was left?" The value of the easiest problem is 2.7 and that of the hardest 9.4. The easiest problem is correctly solved by more than half of the third grade children while the hardest problem is solved by but little more than half of the eighth grade children. A problem is located at each step of the scale and the steps are 0.2 or 0.3 units apart. Thus we have a difficulty scale ranging by small intervals from a typical third grade problem to a typical eighth grade problem.

This year the scale will be enriched by introducing several problems at each level of difficulty. Probably a second or alternative form for the test now presented will also be formulated.

Geography, history, and grammar. A large amount of material in each of these subjects was submitted for evaluation. In geography there were twelve tests each containing twelve questions; in history there were three tests each containing twelve questions; and in English grammar there were ten tests each containing twelve questions. The following figures indicate the scope of the work as it was done last year. In each subject, however, the writer had previously submitted the same questions to large numbers of children.

	Geography	History	Grammar
Number of superintendents and principals approached	40	40	40
Number to whom material was sent	32	17	31
Number who returned material	25	14	29
Number of questions submitted	144	36	120
Number of children participating	9,575	4,066	11,000 (approx.)

The testing in these subjects has not yet eventuated in the construction of scales or standardized tests. Much more material will have to be received before such an outcome can be expected. The subjects do not themselves yield to arrangement in scale form as readily as do some other subjects. It will be the task of the Bureau this year to round out the body of material in hand with reference to these subjects and to issue scales in them if it appears practicable to do so.

Spelling. Using material which the writer had previously obtained for the most part in states other than Illinois, an extension of the Ayres Spelling Scale was devised. This was done with the very generous permission of Dr. Ayres. Five hundred five new words were added to the thousand of which the Ayres Scale consists. These words tended to be more difficult than those of the original scale. It was felt that this was fortunate because admirable as the Ayres Scale is, it is less available for use in upper grades than it is in middle and lower grades because of the fewness of the words at the high end. Since most of the new words fall into the columns at the upper end of the scale, the resulting extension, merely as an instrument for measuring spelling ability, is more satisfactory.

It is to be understood, however, that the Extended Ayres Scale no longer constitutes a fundamental vocabulary, scientifically determined. The added words were not obtained, as were the original words, from tabulations of frequency of use. It is felt by the writer that the new words are useful, but he realizes that this is more or less his personal opinioin. They were selected from a much larger list secured from the agreements among spelling books. It is, therefore, impossible to suppose that these 505 new words are the next most common words after the thousand contained in the Ayres Scale.

In constructing this extension of the Ayres Scale, the original scale was not disturbed. All the new words are printed in the appropriate columns under the old words and in italics. A person, therefore, who secures the "Extension" also has the original Ayres Scale intact.

MEMORY WORK IN THE GRADES

The problem. Early last year, Superintendent H. Q. Hoskinson, of Pinkneyvillle, Ill., wrote the Bureau requesting information as to standards for memory work in the elementary school. He wished to know what poems were suitable material for memorizing and also how much memory work ought reasonably to be required in each grade.

Upon investigation it appeared that there was no literature bearing on this subject. This is rather remarkable since it is quite common to require certain passages to be memorized by children and one would suppose that the selection of these passages and the determination of the amount of memory work which ought to be imposed would have engaged the attention of investigators.

The method. Convinced therefore of the value of more information bearing on the questions raised by Superintendent Hoskinson, we decided to examine the courses of study of elementary schools with a view to finding out what the practice is in selecting material to be memorized. All the printed courses of study, amounting to about two hundred, which the Bureau had on hand were examined to find out what provision was made for this sort of work. Most of the courses of study were silent on the matter, but fifty of them were found to embody usable material. It is at once evident that most of the literary material prescribed or suggested to be memorized consisted of poetry. For the time being, therefore, the study was made without reference to prose selections.

It was found that three hundred eighty-six poems were mentioned five or more times in the fifty courses of study. A poem was often mentioned more than once in the same course of study. A few poems, for example, were listed for memory work in more than one grade. A poem so mentioned was evidently thought of as sufficiently important to be brought to the attention of the teachers several times. The method, therefore, of counting each mention as often as it occurred rather than of counting the courses of study in which the mentioning of a poem took place seemed to be justified. Nevertheless the method of counting each course of study once was also used as a check on the method of frequency of mention. The results were not materially different..

The ten most frequently mentioned poems. The poem most frequently mentioned for memorizing, was "America." In fact patriotic poems tended to be mentioned in courses of study throughout successive grades. On the basis, therefore, of frequency of mention these poems stand higher in the list than they would on the basis of the number of courses of study mentioning them.

The following is a list of the ten poems most frequently mentioned.

No.	Author	Title	Times Mentioned	Cities Mentioning	Weighted Value
1	Smith,	America	59	30	207
2	Longfellow,	The Children's Hour	44	39	135
3	Field,	Dutch Lullaby	43	41	136
4	Longfellow,	Village Blacksmith	42	39	134
5	Stevenson,	My Shadow	42	40	130
6	Scott,	"Breathes there a man—"	40	38	139
7	Jackson,	September	41	36	125
8	Key,	Star Spangled Banner,	38	36	135
9	Stevenson,	The Wind	38	32	112
10	Browning,	Songs from "Pippa Passes"	35	30	105

Many interesting facts were disclosed as this study progressed. They formed the basis of a master's dissertation by Miss Velda Bamesberger. The material has all been worked over a second time and a bulletin will shortly be published for general distribution. If there is a demand for it, the Bureau will attempt to publish a collection of the three hundred eighty-six poems which were mentioned five or more times.

Other results of the study. A few interesting and outstanding facts may be listed in concluding this section. (1) Poems were mentioned in the fifty courses of study from one to fifty-nine times. The number of poems mentioned once was one thousand five hundred fifty-five. One poem was listed fifty-nine times (America). (2) The poems were graded according to the agreements between the courses of study, with the result that from forty to sixty poems were assigned to each of the eight elementary grades. (3) Judged by the number of times mentioned, Longfellow is the most popular poet and Stevenson the next most popular. The following are the eight poets each of whom contributed ten or more poems to the list, the numbers in parentheses after each author's name indicating the number of his poems in the list: Longfellow (25), Stevenson (23), Tennyson (17), Sherman (13), Lowell (12), Field (12), Bryant (10), and Wordsworth (10). (4) A record was also taken of the total number of times the poems by each author were mentioned. For example, Longfellow's twenty-five poems were mentioned a total of four hundred fifteen times; Stevenson's twenty-three poems were mentioned a total of three hundred seventy-six times, etc. On this basis of frequency of mention the first eight most popular poets were the same as those mentioned above although in a slightly different order. Longfellow, Stevenson, and Tennyson were first, second, and third, respectively, both on the basis of the number of poems contributed and on the basis of frequency of their mention. (5) Many of the courses of study did not state the quantity of material to be memorized. Those which did so generally gave the requirements in the form of the number of poems to be committed to memory in each grade. This permits teachers to vary greatly the amount of material by selecting long or short poems. The best general statement is that about six poems are being required to be memorized in every grade from the first to the eighth, with a tendency in grades above the fourth to require seven or eight. The fairly constant requirement of six poems indicates that the amount of material to be memorized increases through the grades. This rests upon the fact that there is a tendency for the poems in the upper grades to be longer than in the lower grades.

Miss Bamesberger in her thesis rightly says: "There are grounds for arguing that no material which enters into the course of study should be more carefully selected than that which is chosen to be memorized by the pupils.

"Out of the vast body of literature in English we sift out a small portion for inclusion in readers and collections of prose and verse. The material rejected enormously exceeds in amount that which is utilized. But the selection of material which pupils will be expected to make their permanent possession must be even more rigidly defined than the material we put into our readers. Thus the body of verse which we are willing to suggest or require to be memorized should represent the very best that our literature affords."

INTELLIGENCE TESTING

The school year 1918-19 was characterized by the derivation of several group intelligence tests for school use. The widespread interest in determining the mentality of school children and the insistance of school people upon some device for doing this at an expenditure of time and money within available resources was at the bottom of the development of these types of tests. The feeling of school people about this matter was manifested in the numerous requests received by the Bureau for advice, suggestions, and material with reference to the intelligence testing of children.

Tests used. Toward the end of the year and under the direction of acting assistant director Charles E. Holley, several of the more important newly devised group intelligence tests were submitted to the children of the Champaign Public Schools. The principal object in doing this was to find out how valuable each of these tests was. The following tests were used: (1) Virginia Delta I, devised by Professor M. E. Haggerty for the use of the Virginia Education Commission; (2) Whipple's Group Test for Grammar Grades, being a series of exercises selected from a larger number used in connection with the gifted class operated experimentally at Urbana in 1916-17; (3) Theisen's Classification Test, being a group of exercises based largely on the army mental test used in the examination of recruits; (4) Otis' Group Intelligence Scale—a series of ten tests devised by Lieutenant Arthur S. Otis and published by the World Book Company; (5) The Primer Scale, devised by Luella M. Pressey; (6) The Picture Completion Test, devised by Dr. C. E. Holley; and (7) The Sentence Vocabulary Test, also devised by Dr. Holley. A total of nearly eight thousand of these tests was used. They were not only furnished by the Bureau but were administered. scored, and interpreted by the officers and employees of the Bureau.* Only a few children were tested with Whipple's Group Test for Grammar Grades and with Holley's Picture Completion Test. Whipple's test required a long time both to give and to score.

Results. The following table shows for the remaining five tests the average scores secured in the grades to which the tests were submitted.

TABLE I SHOWING AVERAGE SCORES OF CERTAIN INTELLIGENCE TESTS FOR CHAMPAIGN PUBLIC SCHOOLS.

TEST	GRADE											
	I	II	III	IV	V	VII	VI	VIII	IX	X	XI	XII
Otis Scale	—	—	—	—	—	98	113	128	133	138	145	149
Classification	—	—	—	—	60	84	98	108	118	125	134	141
Primer Scale	43	56	64	—	—	—	—	—	—	—	—	—
Virginia Delta	—	—	48	69	83	103	112	117	—	—	—	—
Sentence Vocabulary	—	—	17	29	34	41	47	52	57	61	64	66

Evaluation of intelligence tests. In his report Dr. Holley has examined these tests with respect to the following points:

I. The grade to which the test may be suitably offered. Other things being equal, that test is most valuable which gives significant results for a large number of grades.

II. Ease of administration. This means not only the care with which the directions have been worked out and their simplicity, but also the length of time required for giving the test.

---

*The only exception to this occurred in the case of Dr. Holley's Vocabulary test which was administered by the teachers of the respective classes under the close supervision of Dr. Holley.

III. The amount of time required to rate the papers.

IV. The relation between scholarship on the one hand and scores in terms of the scale on the other hand. It is held that fairly high correlation ought to exist between the intelligence and the scholarship of children. Therefore, an intelligence test whose scores do not correlate rather closely with the scholarship ratings made by teachers is presumably an inferior test.

V. Scales which do not reveal sufficiently large differences between the average scores of successive grades manifestly cannot discriminate between the different abilities presented by the children in those grades. Moreover the differences between average scores for successive grades should be reliable differences. Computations of reliability were made as a part of the analysis of the results of these tests.

VI. Finally, the scales may be considered with reference to their component parts. Each one consists of a large number of items, and in most cases these items are grouped so as to bring together those of the same general nature. It is possible that some of the subordinate tests of which the scales are composed may contribute little or nothing to the total score. It may be, for example, that the scores in an analogies test and scores in a similarities test mean much the same thing. It is possible, too, that some of these subordinate tests show inversions as between successive grades. With a view of ascertaining the extent to which these conditions might exist, the analysis of all the scales was made by examining the scores in the subordinate tests. The following table, taken with some modifications, from Dr. Holley's report, indicates the conclusions he reached on these matters.

TABLE II, GENERAL COMPARISONS OF THE SIX INTELLIGENCE TESTS

	1 Scale	2 Range of Suitability (Grades)	3 Time Required for Giving	4 Papers Scored per Hour[b]	5 Correlation with Scholarship	6 Reliability of Total Scores between Successive Grades	7 Individual Tests
a	Otis	6—12	70	13	0.42	poor	Erratic and poorly balanced
b	Classification	5—12	50	15	0.50	poor	Regular but poorly balanced
c	Virginia	3—8	30	20	0.59	good	Regular but poorly balanced
d	Group	6—8	80	6	*	*	Regular and well-balanced
e	Primer	1—3	25	35	0.34	fair	Regular and well-balanced
f	Vocabulary	3—12	20—40[a]	40	0.47	good	Regular and well-balanced

[a] No time limit. Requires from 40 minutes in grade III to 20 minutes in grade XII.
[b] According to our office experience.
*Data not available.

It was the conclusion derived from our experimentation that the present intelligence scales are susceptable of much improvement. Some of them are erratic and poorly balanced In some cases the reliability of the scores is not as high as it should be or as it can be made. Considerable work may be done to advantage in simplifying the administration and scoring; while nearly all of the tests can probably be improved by making every item and every group of items in them contribute to the significance of the total score.

It is the conclusion of Dr. Holley that the best scales for intermediate and grammar grades are the Virginia Delta I, and the Vocabulary Scale. "A combination of these," he says, "seems to offer the best measuring instrument." The Classification Test seems to be the best for the high school grades. The Primer Scale is well organized from the point of view of administration, scoring, and balance of parts. It does not, however, appear to be of highly discriminating or direct value.

An important beginning has been made in the devising of group intelligence scales. It is not to be wondered at that trials made with these scales reveal discrepancies and difficulties. In spite of these there is every reason to believe that successful group scales will be made. If no greater improvement is recorded in this respect than has been shown in the development of tests of school subjects, we shall shortly have instruments for measuring intelligence that will embody all the essentials of such tests.

THESES OF GRADUATE STUDENTS

A considerable part of the research work in education is handled by graduate students under the direction of faculty members. The results of these investigations, although they are written in the form of these or dissertations, are for the most part never published. This is particularly true of dissertations for the Master's degree.

Nevertheless, it is certain that most Doctors' theses and many Masters' theses contain valuable data on significant questions. There should be some way in which these reports can be made available in summarized or bibliographical form so that although we may not have general access to the material, we can at least learn what it is and where it may be found.

It seemed so well worth while to do this, that toward the end of last year the heads of departments of education and of schools and colleges of education throughout the country were asked to submit the authors and titles of the dissertations for the Masters' and Doctors' degrees in Education which had been granted since January first 1917. At this writing almost all colleges and universities have answered this request and an early report of the graduate study in education lately conducted in our higher institutions of learning will be forthcoming.

TYPE LESSONS

A project was proposed last year to bring out a bulletin or series of bulletins containing stenographic reports of type lessons. For example, several good inductive lessons dealing with different subjects taught to different grades might be obtained and reported just as they were taught. Editorial comment justifying and criticising certain procedures would no roubt give value to the report. Perhaps some lessons embodying faulty methods might be given with a discussion of why they are considered faulty. Certain other data concerning the recitation besides the verbatim report might also be included. For example, a description of what was done and the manner in which it was done. Notes of this sort might be inserted after the manner of stage directions. The report might well be accompanied by a seating plan of the room and a statement of the distribution of opportunity to recite.

It was the belief of the writer that a good book of this kind would be helpful both to students in training and to teachers in actual service. A teacher of students in training or a supervisor of teachers in service, may indeed seek to impart the principles applicable to different types of lessons without securing the intelligent practice of such principles. Observation of good teaching would of course, be a better means of securing this end, but actual teaching of the kind that is wanted is seldom available. Moreover, the presence of a printed report makes the discussion of the teaching easier than it is when observed teaching must be remembered after the lapse of perhaps a considerable interval.

At any rate, it was thought that if the publication of such lessons supported by notes and critical apparatus and with lesson plans and charts of rooms could be printed and distributed it might be helpful. With this notion in mind an inquiry was made of fifty-five school people in Illinois, twenty-five of whom responded. All said that they thought the project of publishing such lessons would be helpful, and all but one said he could use such a publication personally with his teachers or students. That one said "Probably."

A number of educational people outside of Illinois were also addressed on the question of whether this project was worth while or not. As in the case of the Illinois people, most of them thought the task ought to be undertaken. Some of the correspondents wrote enthusiastically about it.

Largely for lack of sufficient clerical and stenographic help, however, the Bureau has not yet begun actual work on this project. As one looks at it more closely, it evidently involves no small amount of expense. At least two stenographers would have to be present at each recitation; one would record the words of the teacher and pupils in the usual manner. The other would probably have to work with a large chart divided into squares representing the seats in the room, and would record the words of the pupils in the squares corresponding to their seats. This or some other device would have to be adopted to indicate the identity of the pupils. A single stenographer sitting in a room full of unfamiliar pupils could not be expected to get what was said and identify the row and seat number of the child who said it. An additional observer would probably have to be provided to take account of such things as attitude, facial expression, gestures, illustrative material, black-board work, notes on the text book used, the extent to which the attention of children was held, and the way in which the class-room management proceeded (distribution of material, children passing to the board, etc.) The same observer would indicate the elapsed time for each portion of the recitation period.

Out of all these observations could be drawn up a report which would do something more than is usually attempted in the verbatim reporting of lessons. As a matter of fact—and as some of our correspondents indicated—these reports tend to be rather barren. We think, however, that if they are embelished not only in the ways we have indicated but also by editorial comment and suggestion, they will prove helpful to the student teachers and also to teachers in service. This project will be resumed as soon as we have the competent stenographic and clerical force to handle it.

CHAPTER III. PROPOSED PROJECTS

The work of the Bureau for the current year naturally includes the continuation of certain projects begun last year. The reader is referred to Chapter II for a statement concerning these projects. In addition, however, to the continuance of projects already initiated, the Bureau has in contemplation the starting of certain new projects; and a statement concerning these projects will form the material of this chapter.

Project I. Standardization of educational tests. Before an author publishes a test he submits the material for trial and after eliminating some of it as unsuitable and modifying some of the items in the light of his experience, he brings out the test. He has usually obtained some tentative standards, but the limitation on his time and money prevent his carrying this phase of his work to a conclusion. There is, therefore, no test which does not require after its publication a thorough trial in order to set up norms of performance in terms of it. Many of the older tests have passed through this period and may be said to be satisfactorily standardized.

During the past year certain tests were devised by the Bureau of Educational Research. Certain other tests were taken over for publication and distribution. All of the tests by the Bureau and some of those thus acquired for distribution still stand in need of complete standardization.

The resources of the Bureau are not such that it is able to publish an experimental edition of these tests and distribute it gratis to those who can co-operate in standardization. This is the usual condition. In submitting this project, therefore, of standardization of tests the Bureau is following the customary course. The only difference, however, is that the Bureau is distributing an unusually large number of tests and can therefore offer opportunities for standardizing a rather large variety of different measuring instruments.

On October 6, 1919, Dr. Monroe of the Bureau sent out a circular letter to all the superintendents and school principals who in response to our inquiry last year indicated that they were interested in the study of experimental education and that they were willing to co-operate with each other and with the Bureau in carrying forward such lines of work. In that letter Dr. Monroe offered the services of the Bureau in the tabulation of all returns received and in sending to participants a summary of results.

The following tests were suggested as those which the Bureau desired further to standardize:

1. Buckingham's Scale for Problems in Arithmetic.
2. Monroe's Standardized Reasoning Test in Arithmetic.

3. Monroe's Diagnostic Test in Arithmetic.
4. Harlan's Test of Information in American History.
5. Charter's Diagnostic Language and Grammar Test.
6. Willing's Scale for Measuring Written Compositions.
7. Henmon's Latin Test.
8. Monroe's Standardized Silent Reading Test III.
9. Strach's Physics Tests.
10. Holley's Picture Completion Test for Primary Grades.
11. Holley's Sentence Vocabulary Scale.

It was suggested that those who are interested in participating in this project order material so as to be able to give it during the month of November, 1919, preferably as near the fifteenth of the month as practicable.

Project II. Collection of information about the usefulness of tests. A systematic attempt will be made to obtain certain important facts about all the tests distributed by the Bureau. This project is undertaken in order that we may make more rational recommendation to school people. The real value of a test only becomes apparent through its use, and the Bureau will therefore have to rely upon obtaining information from those who purchase the test material and use it in their schools.

Accordingly, with every package of test material sent out by the Bureau, a questionnaire is included in which the recipient is asked to report on the following topics:

1. The time required in scoring the papers.
2. The time required for tabulating the scores and calculating class medians or averages.
3. Any difficulty experienced in understanding directions, either for scoring or for tabulating.
4. The notation of any errors or weaknesses discovered in the test.
5. Suggestions for modification of the test.
6. A statement of how the test has been used to improve instruction.

This last point is the one in which the Bureau is particularly interested. It is evident that unless the tests in some way function to improve instruction their usefulness is questionable.

This material will be gathered throughout the current year and ought to form the basis of an instructive bulletin to be published at the close of the year or at the beginning of next year.

Project III. Derivation of new test material. The Bureau announces that it will undertake the derivation of a new algebra test this year. Rugg and Clark's Standardized tests in algebra are being extensively used but are expensive and are becoming more so on account of increased cost of printing. These tests also require a rather large expenditure of time on the part of both the pupils and teachers. It seems wise, therefore, to attempt the derivation of a simpler and less expensive test.

Supplementary to Buckingham's Scale for Problems in Arithmetic the evaluation of new problems will be undertaken with a view of enlarging the scale and providing alternative forms.

The Bureau also has in contemplation the construction of a "continuous silent reading test." Further announcement concerning this project will be made later.

Superintendents and principals have already been approached on the question of deriving test material in algebra and in arithmetic problems and have generously responded.

Project IV. Degree of equivalence of the different forms of Monroe's Silent Reading Test. The three forms of Monroe's Standardized Silent Reading Test were constructed so that they would be exactly equivalent. Since their publication it appears that the intention of the author was realized only approximately. The different forms are now being given to several thousand school children in such a way that the degree of non-equivalence can be ascertained. In case a material difference between the forms is found to exist, it will then be possible to give definite directions for comparing scores obtained from different forms.

Project V. Analysis of arithmetical abilities and study of pupils' errors. Fragmentary analyses of the field of arithmetical operations have revealed that there are many arithmetical abilities instead of a single ability. Or to express the same idea in another way, there are many types of arithmetical examples. In close connection with these types of examples are the errors which pupils make. For each type of example there is the possibility of one or more errors peculiar to the type.

The Bureau is undertaking to summarize the work already done on these two related topics and to supplement this by original research where it is needed.

A complete analysis of the operations of arithmetic to determine types of examples is a necessary prerequisite for complete diagnosis. Until we know the types of examples, we cannot ascertain the types of errors which pupils make.

Project VI. Determination of validity and reliability of educational tests. For many years workers in the field of educational measurements have been devising tests until we now have several tests for most of the subjects in the elementary school. In this work little attention has been given to questions of the validity and the reliability of the measures which the tests yield. The availability of a number of tests for several of the elementary school subjects makes urgent a scientific determination of the validity and reliability of the respective tests in order that one may have information on which to base a rational selection.

During the past decade the number of educational tests has increased very rapidly and the use of these tests in public schools has also grown by leaps and bounds. The period has been one of remarkable development. However, a survey of these activities reveals the fact that for the most part we are still using without much refinement the same methods of deriving tests as at the beginning of the period. This perhaps has been necessary in the very rapid development, but the time has now arrived when increasing attention should be given to the refinement of our measuring instruments. This is another reason for the scientific study of the validity and reliability of educational tests because out of such studies there will come helpful information concerning possible refinements. During the present year the efforts of the Bureau in respect to this project will be confined for the most part to silent reading. It is hoped in another year to extend it to other fields.

Project VII. Efficiency of the departmental teaching of mathematics. This topic was proposed by the mathematics section of the High School Conference. The question at issue is whether the departmental teaching of mathematics is more effective than the teaching of that subject in the regular grade organization. There does not appear to be any adequate published data on the subject. Accordingly, the Bureau is taking up the matter with a number of superintendents and principals in whose schools mathematics is taught as a departmental subject and with others in whose schools mathematics is taught by the grade teachers. The first group of schools will constitute the experimental group and the second the control group.

Obviously a single testing of the performance of children in the departmental schools on the one hand and in the non-departmental schools on the other hand would yield no reliable conclusion as to the efficiency of the teaching in the two systems. It has been decided to limit our study to the seventh and eighth grades. It is clear that children in these grades have received their instruction in mathematics largely in a grade organization. What they know, therefore, and what they can do whether they are now in departmental or non-departmental schools will in either case be largely the product of teaching done in the same type of organization—namely, the grade organization.

We shall, therefore, offer two parallel tests in arithmetic, and we plan to do this at the beginning and at the end of the second semester of work during the current school year. Between the initial test offered in February and the final test offered in June, departmental teaching will be carried on in the experimental schools and grade teaching in the control schools. The difference between the performances of pupils on the initial and final tests will, in a general way, be the product of instruction within the period. If the improvement is greater in the departmental schools than it is in the grade schools, we shall have one item of evidence as to the superiority of departmental teaching.

There are, however, a number of variable factors which may enter to modify or even invalidate a consclusion based merely upon the relative amount of improvement between the initial and final tests. These variable factors—or at least the most important ones—must either be rendered constant or allowed for. In order to allow for them we must know their amount for the different classes involved in the experiment. These factors relate for the most part to the teacher, the pupil, and the course of study. There are, it is true, some physical factors relating to the schoolroom, the equipment, and supplies. The effect, however, of factors of this sort on the product of instruction over a short period is probably small—at least what evidence we possess seems to indicate that this true.

It is felt, therefore, that if we can take account of differences among the teachers and pupils participating in the experiment and of differences in topics taught, we should go a long way toward refining the gross measures of efficiency of teaching derived from the progress made between the initial and final tests. Accordingly, we propose to secure data concerning the training and experience of the teachers involved in the experiment and a rating of their teaching ability by means of a score card. We propose also to give an intelligence test of a simple kind to all the pupils in both the experimental and control schools who participate in the test. Finally, we shall attempt to sharpen our judgement by obtaining a daily statement from the teachers of the amount of time spent on the most important topics in arithmetic.

Project VIII. The amount of time spent on topics in arithmetic. To some extent this project enters into the one we have just described but it will include a large number of schools not involved in the experiment to determine the efficiency of departmental teaching. Moreover, it will be extended to include grades below the seventh and eighth. The plan is to furnish teachers with forms on each one of which may be kept the data for one month at a time. A careful analysis has been made of the topics of arithmetic and provision is made for a column for each day of the month. In these columns the teachers will enter the number of minutes of school time devoted to each topic into which the course in arithmetic has been analyzed.

We do not know how much time is spent in teaching decimals, how much time is spent in reviewing the fundamental operations, how much time is devoted to the teaching of percentage and its application. To be sure we have the statements of courses of study but every one knows that courses of study are not strictly followed, and that whether the course of study mentions reviews or not, every successful teacher is obliged to spend a large amount of time reviewing work of previous grades. This project will be an attempt to find out the total amount of time actually devoted to the various topics of arithmetic throughout the grades concerned.

Project IX. The supply of trained teachers and the demand for them. In connection with the committee on the training of teachers of the High School Conference the Bureau will probably undertake this year a statistical study of the number of new teachers annually required in Illinois for each type of school and for special subjects. With this it is proposed to relate the supply of such teachers furnished by the teacher-training institutions of the state. It is pointed out that we do not know, except in a vague way, how many teachers of each of the different types are needed by the schools. We know that the teaching force as a whole amounts to a certain definite number. We may infer, from our general knowledge as to the rate at which teachers drop out of service, the total number of new teachers annually required. We assume, and probably correctly, that the vacancies are filled, but from what source the teachers come, and to what extent they are adequately trained we do not know. We do not know, for example, how many of the vacancies must be filled by teachers of little or no training because teachers of sufficient training are not available. Again, we do not know to what extent certain types of teaching are over-supplied so that those trained for a given type of service are obliged to teach in other lines in order to obtain positions.

It is not too unreasonable to expect that the state should organize its teacher-training establishment so as to take care of the demand for trained teachers however large it may be in whatever line the demand may exist. Moreover, it is right that the State should know to what extent the teachers trained at its expense render service to the schools of the state after receiving the training.

Project X. Supplementary study of memory work. It will be recalled that one of the projects undertaken last year was the determination of standards for memorizing poetry in the grades. Lists of poems agreed upon by the makers of the courses of study, the grouping of these poems into grade lists, and the determination of the amount of memory material usually required—these constituted the objects of the study as it was undertaken last year. This year we are extending the study to include prose selections and proverbs—in fact to cover all the memory work required of children in the elementary school. It is probable that a bulletin on this topic will be published shortly.

Project XI. The selection of text books. Mr. A. E. Capps, instructor and graduate student in education, is co-operating with the Bureau in the study of this topic. The project will involve collecting evidence on the methods now employed in selecting textbooks and the derivation of score cards for judging text books in several of the most important elementary school subjects. In connection with this project and for the broader uses which it may have in every direction, the Bureau is building up a text book library. We are now collecting text books in reading, spelling, geography, and history. These books will be of great value to us in our work. We shall also be able to serve the purposes of the publishers by placing these books on exhibit and making them known to students, faculty members, and school men and women who may use of the facilities of the Bureau.

Project XII. The extension of Monroe's Silent Reading Test III for use with college students. In co-operation with Dr. C. W. Stone of the Department of Education, the Bureau is preparing an extension of Monroe Standardized Silent Reading Test III, which is designed for use in the high school, so that it may be more suitable for use with college students. This extension involves the evaluation of a sufficient number of exercises for both forms of the test so that practically no college students will complete the test within five minutes.

Project XIII. Superintendents' conference. Last year Dean W. W. Charters proposed the establishment of an annual conference of superintendents to be held at the University as one of the activities of the College of Education. The promoting of this project was placed in the hands of the Director of the Bureau; and he communicated with a large number of superintendents in the state with a view of finding out whether or not a superintendents' conference was wanted and would be supported. The evidence from this inquiry seems to indicate that there is a strong demand for a conference to be organized under the auspices of the College of Education and that a large number of superintendents will be likely to support such a conference by attending and participating in its work.

On October 29th and 30th, 1919, the State School Board Association and the City Superintendents' Association met at the University of Illinois. At that time the Director of the Bureau presented before a joint meeting of the Associations the proposal that the University through the College of Education is making. The Superintendents' Association at its business meeting appointed a committee to consider the action which the association ought to take with reference to this proposal and to report at a meeting to be held in connection with the State Teachers' Association during the Christmas Holidays.

If the superintendents though their association or otherwise continue to manifest an interest in the organization of a conference, it will be the work of the Bureau this year to plan such a conference, arrange a program, engage speakers, and print the proceedings. According to the present indications, the proposed ors of educational research, and other persons engaged actively conference, if it is held, will meet during the same week that the High School Conference meets. The first meeting would, then, be held in November, 1920.

XIV. The Journal of Educational Research. Beginning in January 1920, the University of Illinois through the Bureau of Educational Research will bring out a new magazine called the Journal of Educational Research. This journal, although it is being promoted by the University of Illinois, will be by no means narrow in its scope. That this is true is clearly shown by the Board of Associate Editors: Leonard P. Ayres, of the Russell Sage Foundation, in charge of child accounting; W. W. Charters, of the Carnegie Institute of Technology, in charge of Curriculum Analysis; S. A. Courtis of the Detroit Public Schools, in charge of Supervision and Bureau Organization; Walter S. Monroe of the University of Illinois, in charge of Educational Tests and Measurements; George D. Strayer, of Teacher's College, Columbia University, in charge of Educational Finance, Buildings and Building Programs, Records and Reports; and Lewis M. Terman of Leland Stanford Jr. University, in charge of Intelligence Tests.

In addition to the associate editors, there is also a large Board of Contributing Editors consisting of a number of direct- in research work.

While the journal will not entirely avoid theoretical and technical articles, it will especially emphasize the practical application of the results and methods of educational research. To the actual worker in the schools it seems that investigation of educational questions has been overloaded with theory. Of course it is a fact that the practical bearings and implications of these theoretical statements are very numerous, and it is to the task of showing these bearings and making these implications explicit that the Journal of Educational Research will be devoted. How the teacher and the supervisor can utilize the results of the experimentalist, what methods may be be transferred from the laboratory to the schoolroom and the school office—these are the questions with which the new journal will be especially concerned.

No small amount of the energies of the director of the Bureau—who is also editor in chief of the journal—and no small amount of the energies of the employees of the Bureau will be directed to furnishing the editorial supervision of this magazine. It will be their endeavor to make it the highest and best expression of the service which educational investigators can render to the public schools of the nation.

CHAPTER IV. THE WORK OF THE LIBRARY DIVISION.

Need for a library division. As our relations with the public schools of the State have become closer, the tendency for superintendents and principals to refer matters to us for decision and to seek information from us has increased. No small part of our time is at present occupied in supplying information to school people. This type of service, therefore, while it has already assumed rather large proportions may be expected to absorb an even greater portion of the time and energy of the employees of the Bureau during the present year.

As has been indicated elsewhere, an extensive correspondence is carried on with users of test material. The questions that are asked of us in connection with this work are very numerous and require a great deal of thought on the part of both the administrative and clerical staff of the Bureau.

It is not, however, of this work that we wish to speak in this chapter. The need for advice, information, and evidence on numerous topics concerned with the work of the schools has necessitated our setting up in the Bureau a rather extensive and somewhat peculiar library. There are very few bound books in this library. By the time educational material has reached the stage of being put into such books, it has become easily available in general libraries. The need, therefore, for an extensive collection of bound volumes has not been felt.

Organization and routine. The library of the Bureau consists of pamphlets, bulletins, reports, and periodicals—in short, of that type of material known to librarians as ephemeral. The kind of material contained in these volumes is various but a large part of it may be classified as consisting of courses of study, school laws, annual reports of boards of education, reports of surveys, proceedings of learned societies, publications of educational foundations, and bulletins of research departments and of other educational organizations. A practically complete file of the bulletins of the Federal Bureau of Education is available. The material in the library now consists of about 3050 titles, to which should be added the fifty-nine periodicals which the Bureau regularly receives.

Perhaps there is nothing remarkable about the material of which the library consists. Many other libraries doubtless have much the same collection. But in most cases of which the writer has information the material is practically worthless because no one knows what it contains. It has, therefore, been the problem of the Bureau to make this material as useful as it is capable of being. As volumes and pamphlets come in, they are first accessioned, classified, and catalogued. This involves making of author and title cards as well as subject cards. The material—bulletins, reports, etc.—is itself arranged according to a fixed classification which has reference to the issuing body and the purpose of the publication.

But this is not all. Each article is read and analyzed according to a classification of educational activities, prepared for the purpose. For each topic treated in the volume in question, a card is made with page references. Not infrequently from one to two hundred topical cards are thus made with reference to a single report.

In addition to this topical index, all references to the use of tests are noted and in each case two cards are made, one for the test and the other for the city or geographical division where the test was used. For example if it is found from the report of the Board of Education of Des Moines, Iowa, that the Courtis Arithmetic Test, Series B, has been used, two cards are made. One says "Courtis Standard Research Tests in Arithmetic, Series B, used at Des Moines, Iowa," followed by the title of the report and the page reference. The other card will say, "Des Moines, Iowa, used Courtis Standard Research Test in Arithmetic, Series B," again followed by the name of the report and the page reference. Thus, as far as published reports are concerned, a person who consults our files may find out where each educational or mental test has been given. He may also learn what tests have been used in each city.

Bibligraphical service. But this organization of the library material is only justified by the use that can be made of it. We receive, as has been said, many requests for information from the school people of the state. In responding to these requests the resources of the Bureau Library are drawn on through the use of the index and filing cards mentioned above. In addition, however, to the utilization of the material belonging to the Bureau Library, we also use (especially for material published in bound volumes) the larger resources of the University Library.

The following are some of the bibliographies which have been prepared by the library division of the Bureau. We use the term bibliographies because it covers perhaps more appropriately than any other single word the type of response to inquiries which we are listing below. Not infrequently the bibliography is largely annoted and in some instances our response consists of a series of suggestions drawn from a large numbers of sources which are not always indicated. The reader's attention is called to the nature of these bibliographies. Copies of them are on file in the Bureau; and if the reader of this bulletin is interested, he may obtain the same information by addressing the Bureau.

LIST OF BIBLIOGRAPHIES PREPARED BY THE BUREAU

1. Select bibliography on the democratization of the school.

2. Select bibliography on promotions, percentages or failures and withdrawals.
3. Select bibliography on mental testing.
4. A list of school reports containing material on the junior high school·
5. Select bibliography on the judgement of textbooks.
6. Select bibliography on programs for parents-teachers associations.
7. Games for recess. A list of books containing suggestions,
8. Vocational guidance. A bibliography.
9. Supplementary material for the teaching of Caesar.
10. Select bibliography on the question of a Federal Department of Education.
11. Books and pamphlets explaining the grouping of grades; 8-4 plan, 6-6 plan, and the 6-3-3 plan.
12. A brief bibliography on "How to deal with pupils below the average."
13. A brief bibliography on "How to deal with pupils above the average."
14. What experiments have been tried in segregation with respect to ability.
15· A brief bibliography on curriculum reconstruction due to the war.
16. Suggestive supplementary material for ancient history.
17. A bibliography on making programs for high school.
18. Community activities centering in and about the high school.
19. Lists of books suitable for children in different grades.
20. A brief biography of the rating of teachers.
21. A biography on the derivation and the use of score cards.
22. Data concerning an all-year-round school. Based on school reports.
23. Bibliography on project teaching in the grades.
24. A bibliography on a point system for extra-classroom activities in high schools.
25. A bibliography on marking systems·
26. Brief bibliography on the departmental plan in the grades.
27. Courses of study containing problems in geography.
28. Suggestions on supplementary material for English history.
29. Suggestions on supplementary material for geometry.
30. Suggestions on teaching high school Latin.
31. Suggestions on supplementary material for high school history.
32. Suggestions on a curriculum for the junior high school in a city having a population of five or ten thousand.

33. A brief bibliography and notes from state laws relating to compulsory education.

34. A bibliography on the relation between quality of teaching and teaching experience.

35. A brief bibliography on building and equipment.

36. Bibliography on credit for quality.

37. The public school as an agency in the growth and development of a democracy. A bibliography.

38. Suggestive reading on vocational guidance.

39. An outline for the study of tests and measurements for high school use.

40. A bibliography on the desirable outcomes from the teaching of history.

41. When should the introduction of formal arithmetic in the grades take place.

42. A bibliography on the project method and socialized recitation in the high school.

An elementary school textbook library. We have lately begun to collect a textbook library. People engaged in school work often write to us about textbooks and call upon us to discuss textbooks. If we had immediately at hand a competent textbook library we could use it to advantage. Moreover, some of the research problems with which the Bureau is dealing and will deal in the future, involve the use of textbooks. So far we have only thought to obtain textbooks in elementary school subjects. A good library of high school textbooks is maintained in the office of the High School Visitor. Moreover, we have not hitherto had frequent occasion to consult such books. We have begun by collecting the best currently used textbooks in reading, arithmetic, spelling, geography, and United States history. We now have about four hundred volumes and the number is growing rapidly.

In the service work which the Bureau will do in the future the reference library and the textbook library will play an important part. The reference library will help us to place at the disposal of school people types of material not readily accessible, while the textbook library will enable us to converse with the school people in a common language.

CHAPTER V. MEASURING THE RESULTS OF TEACHING FOR THE PURPOSE OF IMPROVING INSTRUCTION

The use of educational tests should result in an improvement of the instruction. They are, however, not drill exercises nor teaching devices. Pupils will derive little or no benefit from being "exposed" to the tests. In order to understand how educational tests may be used to improve instruction, it is necessary to have in mind certain things concerning their nature.

What standardized tests are. Educational tests differ from examinations such as teachers prepare in three respects: *First,* the questions or exercises have been scientifically selected with respect to the purpose which the test is designed to fulfill. *Second,* the exercises have been arranged according to certain principles of scale construction so as to form a scientific measuring instrument. *Third,* the test, or scale, has been *standardized,* that is, the score which pupils in the several grades, in which the test is given, make, has been determined. Such a score is called a standard score and the test is spoken of as a *standardized test.*

Diagnostic vs. general tests. Each school subject includes a number of abilities which are specific or distinct from each other to a considerable degree. By this statement we mean that in the operations of arithmetic, for example, a pupil may be able to do short-column addition (three or four figures) satisfactorily when he is conspicuously below standard in ability to do long-column addition (thirteen or more figures). In spelling, a pupil may be able to spell correctly certain words and at the same time misspell other words which in general are not more difficult. In fact it has recently been shown that each derived form of a word constitutes a distinct spelling problem.

A diagnostic test is one which furnishes a separate measure of each specific ability in the field of the test or at least of the important abilities. An excellent illustration of a test which does this is Charters' Diagnostic Language Test for Pronouns. Since the field of pronouns is limited, a scientific survey was made to determine what were the common errors in the use of pronouns. The test includes all the errors which were found and gives information whether or not the pupil is able to correct each error. Since the test completely covers the field of pronouns, it gives a complete *diagnosis* of the pupil's language abilities so far as pronouns are concerned.

A *general test* is the opposite of diagnostic test. It yields average or composite measures in which the magnitude of the individual abilities are lost sight of. An excellent example of a general test is the Courtis Supervisory Test in Arithmetic. This is a single test covering the entire field of all the operations with integers. The test gives only the pupil's general or average standing. It tells nothing about his specific weaknesses.

Some tests are partly diagnostic and partly general. This is true of such a test as the Courtis Standard Research Test, Series B. It is diagnostic to the extent that it yields separate measures for the four operations and for integers only, but is general for each operation.

Certain tests are general or diagnostic, depending upon the way in which the facts secured by giving the test are tabulated. Charters' Diagnostic Language Test for Pronouns, referred to above, is diagnostic only when the facts of the tests are arranged to show diagnosis. They may be arranged to show only a general score.

How standardized tests are helpful in improving instruction. Standardized tests render assistance to the teacher in three ways: *First,* since the test has been constructed after a careful analysis and survey of the field, it gives the teacher a list of the things which the pupil should be able to do. This is well illustrated by the Ayres' Spelling Scale which contains the 1,000 most frequently used words, and by Monroe's Diagnostic Tests in Arithmetic which give a list of the significant types of examples in certain fields. *Second,* since the tests are *standardized,* the teacher may know just what scores pupils in general make. In other words, the teacher is given a definite objective aim to strive for in her teaching, an aim which the pupil can understand. The advantage of a definite standard is obvious. *Third,* tests furnish the teacher with information concerning the abilities of her pupils. They point out phases of strength and weakness. With this information at hand she can plan instruction which will be more efficient since it will meet specific needs.

Since diagnostic tests furnish more detailed information than general tests do, it follows that they are more helpful in improving instruction. This point will be referred to again under the head of "How to Choose a Standardized Test."

Service of the Bureau of Educational Research. Because the Bureau of Educational Research believes that standardized tests are effective instruments for improving instruction, certain facilities have been provided for assisting superintendents and teachers in their use.

1. A prerequisite to the use of standardized tests is that they may be easily obtained. In order that the tests listed on the following pages may be readily accessible to the educational public, the Bureau of Educational Research maintains a stock of them and sells them at cost; consequently it is unnecessary to write to the different individual publishers in order to secure the tests recommended for use.

2. A test cannot be conveniently used unless it is accompanied by complete directions for giving it, scoring the test papers, and tabulating the scores. Further, educators must have record sheets adapted to the recording of the scores. Every test listed for sale by the Bureau is provided with the needed accessories. In order to render this service it has been necessary in some cases for the Bureau of Educational Research to arrange directions and devise record sheets because some authors of tests have not provided these aids.

3. The best available standards are either printed on the class record sheet or are given on a separate sheet. In the case of tests that are just being standardized the purchaser can secure the standards as soon as they are available by writing to the Bureau of Educational Research.

4. The standardized tests listed on the following pages are accompanied by two record sheets for each class of 25. One record sheet will be printed on blue paper and marked "duplicate." Those who purchase test materials are requested to make out a duplicate at the same time that the original class record sheet is compiled and return it to the Bureau of Educational Research. The Bureau sells all its test materials at cost and in return for this service cooperation is asked from those who purchase the materials. From these class record sheets summaries will be prepared at the end of the school year and sent to those who have contributed results. By means of these summaries, school people will be able to compare their average or median scores with similar scores of other cities as well as with the state standards.

5. The value of standardized tests is realized through the use that is made of the scores. They must be *interpreted in terms of the needs of the pupils for instruction.* This means doing more than simply ascertaining whether the scores are above standard, at standard, or below standard. It means doing much the same sort of thing the physician does after he has ascertained his patient's pulse rate, temperature, and other symptoms when he prescribes treatment. The physician interpretes or translates the facts or measures of his patient's condition in terms of his needs for treatment. Because the Bureau of Educational Research believes that this step is necessary for the effective use of standardized tests, it is prepared to render assistance in the interpretation of scores. Helpful suggestions will be found in the literature suggested below as references for the tests, and in general manuals.*

---

*Monroe, Walter S. "Measuring the Results of Teaching," Boston: Houghton-Mifflin Co., 1918.

Starch, Daniel "Educational Measurements," New York: McMillan, 1916.

Monroe, Walter S., De Voss, James C., and Kelly, Frederick J. "Educational Tests and Measurements." Boston: Houghton-Mifflin Co., 1917.

6. The Bureau of Educational Resesarch will carry on investigations in order to improve existing tests and to devise new ones. The cooperation of teachers and superintendents is solicited·

How to choose a standardized test. Several standardized tests are available for most school subjects and for this reason it is necessary to make a choice. The list of standardized tests given on the following pages does not include all available tests. A first or preliminary selection has been made, but it has been thought advisable to list several tests for most of the school subjects. Hence, it is necessary to make another selection from this list. Sufficient information is not yet available to enable one to make a scientific selection; but it will be helpful for the superintendent or teacher to have certain criteria in mind.

1. The use of a standardized test should result in the improvement of instruction, and this is accomplished by means of the information which the test yields concerning the abilities of the pupils. *Diagnostic tests* furnish more information than *general tests.* Hence, if other things are equal, a diagnostic test is to be preferred to a general test, unless the test is being used merely for administrative purposes.

2. All tests do not yield the same kinds of information. *Difficulty tests* show hard the exercises are that the pupils can do. *Rate tests* show how rapidly and how well they can perform the tasks. Other tests show only how well pupils can do certain kinds of exercises. Standardized tests should be selected with reference to the kind of information which is most important for the purpose in mind.

3. An important consideration is that the test should be easy to use. Further, it should not require a large expenditure of time in the giving or in the scoring and tabulating of results. The use of a standardized test is not complete when the scores are recorded on the class record sheet and the class scores calculated. There remain *two* important steps: *First,* the interpretation of the results in terms of pupils needs; and *Second,* the planning of corrective instruction to remedy deficiencies. If a large amount of time is consumed in giving the test, in scoring and in tabulating, the subsequent steps are likely to be omitted.

CHAPTER VI. STANDARDIZED TESTS FOR THE ELEMENTARY SCHOOL

In this and the following chapters will be given in schematic form the most salient facts concerning those tests in school subjects which the Board of Educational Research will carry in stock and will sell at the prices quoted. Many of these instruments of measurement are made up of a series of tests. In these chapters each of these groups of questions or tests is called a test. Thus, the publication known collectively as "The Cleveland Arithmetic Test" consists of 15 subordinate tests; the Courtis Standard Research Test, Series B, is composed of 4 tests, etc.

In many instances authors have published two or more measuring instruments of the same kind and of the same difficulty, but consisting of entirely different material. Each of these publications is called in these chapters a *form* of the test in question. Thus, there are four forms of the Courtis Standard Research Test, Series B· Either of these forms may be used to measure the same sort of ability, and scores obtained through the use of one of these forms are regarded as having the same meaning as scores obtained in the use of either one of the other three forms.

ARITHMETIC

I. Buckingham Scale for Problems in Arithmetic. 1. *Description.* This scale consists of three divisions called Division 1, Division 2, and Division 3 which are designed for different grades. Each division consists of ten problems evaluated in terms of difficulty. Difficulty is understood to mean the difficulty of getting the correct answer. This scale provides material for testing grades three to eight.

2. *Function.* This scale is intended to be used in estimating the ability of children to solve such types of problems as are usually met in courses of study for elementary schools. In spite of emphasis upon the teaching of problems, it appears that the ability of children to read, understand, and successfully deal with the situations presented in verbal problems is rather small. The extent to which this is true in each school may be determined by the use of this scale.

3. *Time Required.* Pupils are to be given all the time they can use profitably. It is suggested, however, that a test should not last more than an hour in any case.

4. Test 1 is for use in grades 3 and 4, Test 2 and grades 5 and 6, and Test 3 in grades 7 and 8. The problems in Test 3 are in most cases, sufficiently difficult for use in the high school.

5. One test is required for each pupil. In ordering specify the number of pupils to be tested in each grade. Other forms will follow.

6. *References.* None are available.

7. *Price*: Including class record sheets and directions, 70 cents per hundred, Postage extra.

II. Cleveland Survey Arithmetic Tests. 1. *Description.* This series of tests on the operations of arithmetic was first used in the Cleveland survey, and since then has been used widely. There are fifteen tests in the series, each composed of different types of examples with integers or common fractions. Only two of the tests are devoted to fractions.

2. *Function.* This series of tests yields measures of the rate and accuracy with which pupils are able to do certain types of examples. The series of tests has been called spiral because the abilities of the pupils are measured on successive levels of difficulty. The series is diagnostic rather than general.

3. *Time Required·* The total working time for giving the series is 22 minutes. The Bureau of Educational Research (which has taken over from the University of Chicago the distribution of the Cleveland Arithmetic Tests) has devised a folder containing directions, standards, answers, and class score sheet which will facilitate the administration and scoring of the tests.

4. The series can be used in grades 3 to 8 inclusive, although some of the tests are too difficult for the pupils in the lower grades.

5. In ordering give the number of pupils to be tested. There is only one form of the test.

6. *References.* Judd, Charles H.: "Measuring the Work of the Public Schools," New York, Russell Sage Foundation, 1917; Counts, G. S.: "Arithmetic Tests and Studies in the Psychology of Arithmetic," Supplementary Educational Monographs, v. 1, no. 4, Chicago, University of Chicago Press, 1917; O'Hern, J. P.: "The Practical Application of Standard Tests in Spelling, Language, and Arithmetic," Elementary School Journal, 18: 662-679, May, 1918.

7. *Price*: Including complete directions and class record sheets, $1·90 per hundred. Postage extra.

III. Courtis Standard Research Tests, Series B. 1. *Description.* These are the well-known Courtis Arithmetic Tests. The series consists of four tests, one on each of the operations with integers.

2. *Function.* This series of tests yields measures of the rate and accuracy with which pupils can do one type of example in each of the four fundamental operations with integers but in each case the type of example has been so chosen as to include all other types. The tests are general in the field of each operation but diagnostic to the extent that information is secured for each fundamental operation.

3. *Time Required.* The total working time for giving the four tests in 26 minutes. The test papers of a class of 40 pupils can be scored and the scores tabulated in two hours.

4. Series B is recommended for use in grades 4 to 8 inclusive.

5. There are four forms of the test. During the school year of 1919-1920 the Bureau of Educational Research will carry in stock only Form 1 and Form 3. In ordering, the form desired should be specified and the numbers of pupils should be given.

6. *References.* The various annual accountings published by Mr· Courtis and his *Manual of Instruction for Giving and Scoring the Courtis Standard Tests*; Monroe, Walter S.: "A Report of the Use of the Courtis Standard Research Tests in Arithmetic in Twenty-four Cities," Emporia, Kansas, Bureau of Educational Measurements and Standards, 1915; "Arithmetic—The Courtis Standard Tests in Boston, 1912-1915. An Appraisal," Boston, Department of Education Investigation and Measurement, 1916; Haggerty, M. E.: "Arithmetic: A Co-operative Study in Educational Measurements," Indiana University Bulletin, Studies, No. 27. University of Pennsylvania Schoolmen's Week Proceedings, p. 97ff, 114ff, 203, April 11-13, 1918.

Price: Including complete directions and class record sheets, $1.00 per hundred. Postage extra.

IV. Monroe Diagnostic Tests in Arithmetic. 1. *Description.* Tests 1 to 11 are on the operations with integers, tests 12 to 16 are on common fractions, and tests 17 to 21 are on decimal fractions.

2. *Function.* As the title indicates, the function of this series of tests is to yield diagnostic measures of arithmetical abilities in terms of rate and accuracy. They do this more completely than do the Cleveland Survey Arithmetic Tests. With the exception of the operations with mixed numbers the diagnosis is practically complete.

3· *Time Required.* The total working time for the 21 tests is 30 minutes. Answers, directions, and record sheets are furnished with the tests so that the time for scoring and tabulating is reduced to a minimum.

4. Tests 1 to 11 can be given in grades 4 to 8, tests 12 to 16 in grades 5 to 8, and tests 17 to 21 in grades 6 to 8.

5. Only one form of the tests is available. In ordering, specify the particular tests desired and give the number of pupils in each grade.

6. *References.* Monroe, Walter S.: "A Series of Diagnostic Tests in Arithmetic," The Elementary School Journal, vol. XIX, no. 8, April, 1919.

7. *Price*: Including complete directions and class record sheets, 70 cents per hundred each part. Postage extra.

V. Monroe Standardized Reasoning Tests in Arithmetic.

1. *Description.* The tests consist of list of problems which have been carefully selected as being representative of the one and two-step problems of our arithmetic text books. The problems are so arranged that the pupil can do all of his work on the test paper.

2. *Function.* These tests are designed to measure the ability of the pupil to reason apart from his ability to perform the operations required in solving a problem. Each pupil receives a score for correct reasoning (correct principle) and another for performing the operations correctly (correct answer). The tests are general in the field of solving one and two-step problems.

3. *Time Required.* For giving the test 25 minutes are required. The scoring of the test papers and the recording of the scores for a class of 25 pupils can be done in one and one-half hours.

4-5. Test I is for grades 4 and 5, Test II is for grades 6 and 7, and Test III is for grade 8. There are two forms. In ordering indicate the number of pupils in each grade, and specify the form desired.

6. *References.* Monroe, Walter S.: "The Derivation of Reasoning Tests in Arithmetic," School and Society, 8:295, 324, September 7, 14, 1918.

7. *Price*: Including complete directions and score sheets, 65 cents per hundred. Postage extra.

VI. Woody Arithmetic Scales, Series B. 1. *Description*. These scales are four in number, one for each of the fundamental operations. Each consists of examples arranged in the order of increasing difficulty.

2. *Function.* The scales are useful in providing a general showing for a class or school. They also furnish considerable information relative to the abilities of the different children tested. The scores obtained from this series may be recorded so as to yield a diagnosis in so far as the types of examples are represented.

3. *Time Required.* Each scale requires 10 minutes for its administration. The time required in scoring is about the same as that required to mark the same number of arithmetic papers.

4. The series can be used in grades 3 to 8 inclusive. They are not so useful in the lower grades, however, because but few examples are offered which the children can do.

5. Only one form of the test is offered. In ordering, give the number of pupils to be tested, and the fundamental operations in which tests are desired.

6. *References.* Woody, Clifford: "Measurements of Some Achievements in Arithmetic," New York, Teachers College, Columbia University, 1916; Anderson, C. J.: "The Use of the

Woody Scales for Diagnostic Purposes," Elementary School Journal, 17:770-781, June, 1918; Monroe, Walter S.: "Experimental and Analytical Study of Woody's Arithmetical Scales, Series A," School and Society, 6:412-420, October 6, 1917; Theissen, W. W. and Fleming, Cecile White: "The Diagnostic Value of the Woody Arithmetic Scales. A Reply," Journal of Educational Psychology, vol. IX, 475-488; 567-580, November and December, 1918; Theissen, W. W.: "A Report on the Use of Some Standard Tests for 1916-1917," Madison, Wisconsin, State Department of Public Instruction, 51-70, 1918. See also Research and Measurement Bulletin No. 2, Pittsburgh, Pa., Janary, 1918.

7. *Price*: Including directions and class score sheet, 50 cents per hundred each part. Postage extra.

GEOGRAPHY

I. Courtis Standard Supervisory Test in Geography.

1. *Description.* This test includes only the states and important cities of the United States. A pupil's knowledge of these is tested by his ability to locate them on the outline map.

2. *Function.* The function of this test has been indicated in its description. Its measures the rate and accuracy with which the pupil can locate the states and 30 important cities.

3. *Time Required.* Six minutes are required for giving the test. It is so arranged that the scoring of the papers and the recording of the scores can be done quickly.

4. This test may be used in any grade after the pupils have studied the United States.

5. The same test is used in all grades and only one form is carried in stock. In ordering, give the number of pupils to be tested.

6. *References.* Courtis, S. A.: Measuring the Effects of Supervision in Geography, "School and Society, 10:61-70, July 19, 1919.

7. *Price*: Including directions and class record sheets, $1.00 per hundred. Postage extra.

II. The Hahn-Lackey Geography Scale. 1. *Description.* This scale consists of 217 carefully selected questions classified according to difficulty and standardized for the several grades. The material covered by the questions is common to six modern texts.

2. *Function.* This scale is not a test. Its function is simply to provide the teacher with a carefully selected list of standardized questions from which she can construct a standardized test.

3. *Time Required.* The time required for giving a list of questions selected from this scale will depend upon the number of questions but is will not be greater than is necessary for an

equal number of questions prepared by the teacher. In fact it will in general be less because these questions have been formulated for the special purpose of measurement. The scoring and recording of scores is facilitated by a score card and record sheet.

4. It may be used in grades 4 to 8.

5. Only one scale is needed for each teacher. One class record sheet is needed for each grade. Give the number of copies of the scale and the number of class record sheets desired.

6. *References.* Lackey, E. E.: "A Scale for Measuring the Ability of Children in Geography," Journal of Educational Psychology, 9: 443-451, October, 1918; Report of the Superintendent of Schools, Board of Education, Lincoln, Nebraska, p. 103, 1915-1917.

7. *Price*: Including complete directions and one class record sheet 10 cents each, extra class record sheets, one cent each. Postage extra.

HANDWRITING

I. Ayres Handwriting Scale, "Gettysburg Edition."

1. *Description.* This is the latest edition of Ayres' handwriting scales. Numerous changes have been introduced which make it a more accurate instrument than the "three slant edition." It uses the first three sentences of the "Gettysburg Oration" as copy material.

2. *Function.* This scale is designed to yield a general measure of the quality of handwriting. The directions for securing a sample of the pupil's handwriting also provide for measuring the rate at which he writes.

3. *Time Required.* For collecting samples of handwriting only two or three minutes are required. The rating of the samples by means of the scale requires more time although they may be rated rapidly. For one inexperienced in using the scale a period of practice is recommended before attempting to make final ratings.

4. The scale may be used in all grades in which handwriting is taught.

5. Only one scale is needed for each teacher. One class record sheet is needed for each class. Give the number of scales and class record sheets desired.

6. *References.* Breed, F. S.: "The Comparative Accuracy of the Ayres' Handwriting Scale, Gettysburg Edition," Elementary School Jouranl, 18:458-463, February, 1918; Report of the Superintendent of Schools, Huron, South Dakota, pp. 60-64, 1916-1917.

7. *Price*: Including one class record sheet, 7 cents each, extra class record sheets, one cent each. Postage extra.

II. Freeman Chart for Diagnosing Faults in Handwriting.

1. *Description.* This scale consists of samples arranged accoring to merit under each of five headings, namely; (1) uniformity of slant, (2) uniformity of alignment, (3) quality of line, (4) letter formation and (5) spacing. Three degrees of each of these characteristics are included in the scale. They are numbered 1, 3, and 5. The intermediate values 2 and 4 may be used.

2. *Function.* As its name implies, this scale is to be used for the purpose of diagnosis. In this respect it is a sharp contrast with both Ayres' Scale and Thorndike's Scale.

3. *Time Required.* The time required for rating samples is of course greater than that required by the Ayres' Scale since each example is rated on five different scales. However, it is recommended that this scale be used only when a sample of handwriting is below standard and one desires to ascertain its particular faults.

4. The scale may be used in all grades in which handwriting is taught.

5. Only one scale is needed for each teacher. Give the number of scales desired.

6. *References.* Freeman, F. N.: "The Teaching of Handwriting," Riverside Educational Monographs, Boston, Houghton-Mifflin Co.; Freeman, F. N.: "Handwriting," Survey of the St. Louis Public Schools, St. Louis Board of Education, 1917; Freeman, F. N.: "Handwriting Test for Use in School Surveys." Elementary School Journal, 16: 299-30, February, 1916; Freeman, F. N.: "An Analytical Scale for Judging Handwriting," Elementary School Journal, 15:432-441, April, 1915.

7. *Price*: Including complete directions, 30 cents each. Postage extra.

III. Gray Standard Score Card for Measuring Handwriting. 1. *Description.* This represents a different type of measuring instrument. It is similar to the score cards which have been used successfully by agriculturists. Nine characteristics of handwriting are considered separately and a specified number of points are allowed for perfection in each. There are four forms of the score card. Form I for individual practice, Form II for actual work in the judgment of handwriting of pupils, Form III class score card for displaying class scores, Form IV a score card for reporting to parents.

2. *Function.* This scale is both diagnostic and general. The total score is a general measure and the scores for separate characteristics are diagnostic measurements.

3. *Time Required.* Exact data are not available on this point, but it is probable that more time will be needed to judge handwriting on this scale than to rate it with any of the other measuring devices.

4. This scale may be used in all grades in which handwriting is taught.

5. One copy each of Form II and of Form IV are needed for each pupil in the class. One copy of Form III is sufficient for the entire class. Only the number of copies of Form I that are to be used for practice need be ordered.

6. *References.* Gray, C. Truman: "Score Card for the Measurement of Handwriting" Bulletin of the University of Texas, No. 37, Austin, Texas, 1915.

7. *Price*: Including complete directions: Forms I and II, 50 cents per hundred; Form III, 10 cents each; Form IV, 30 cents per hundred· Postage extra.

IV. Thorndike Handwriting Scale. 1. *Description.* This scale consists of a series of samples of handwriting arranged in order of general merit.

2. *Function.* This scale is designed to yield a general measure of the quality of handwriting. The directions for securing a sample of a pupil's handwriting provide also for measuring the rate at which he writes.

3. *Time Required.* For collecting samples of handwriting only two or three minutes are required. The rating of the samples by the scale requires more time although they may be rated rapidly. For one inexperienced in using the scale, a period of practice is recommended.

4. This scale may be used in all grades in which handwriting is taught.

5. Only one scale is needed for each teacher. One class record sheet is needed for each grade. Give the number of scales and class record sheets desired.

6. *References.* Thorndike, E. L.: "Handwriting," Teachers College, Columbia University, 1910; Thorndike, E. L.: "Teachers' Estimates of the Quality of Specimens of Handwriting," Teachers College Record, 15:279-291, November, 1914; "A Review of the Rockford Public Schools," Rockford, Illinois, Board of Education, pp· 57-60, 1916; "Annual Report of the Des Moines Public Schools," pp. 80-85, 1915; "Bulletin No. 1 Bureau of Research and Efficiency," Kansas City, Missouri, pp. 34-64, 1916.

7. *Price*: Including complete directions and one class record sheet, 10 cents each, extra class record sheets, one cent cents per hundred. Postage extra.

HISTORY

I. Harlan Test of Information in American History.

1. *Description.* This is an information test in American history. It consists of 10 exercises based on an analytical study of text-books in American history.

2. *Function.* This test gives the teacher a measure of the efficiency with which the information side of American history has been taught.

3. *Time Required.* Most pupils finish the test within 25 minutes. The scoring of the papers can be carried on rapidly because the material is so arranged that the scoring time is reduced to a minimum.

4. This test is best suited for use in the grammar grades. It may be used in lower grades where history has been taught and in some cases in the first year of high school.

5. There is only one form available. One copy is needed for each pupil. Give the number of pupils to be tested.

6. *References.* None available.

7. *Price*: Including complete directions and score sheets, 75 cents per hundred. Postage extra.

II. Sackett Scale in United States History. 1. *Description.* This scale is composed of seven different tests based on carefully selected information in United States history.

2. *Function.* This test gives a general measure of the grasp of the information side of United States history.

3. *Time Required.* This test requires exactly 39 minutes of working time for the pupils. The scoring of papers would require about the same.

4. This scale can be used with pupils who have studied history above the sixth grade. It is difficult enough to test most high school and college students.

5. One copy of the test is needed for each pupil in the class. Give the number of pupils to be tested.

6. *References.* Bell, J. C. and McCollum, D. F.: "A Study of the Attainments of Pupils in United States history," Journal of Educational Psychology, 8:257-274, May 1917.

7. *Price*: Including class record sheets, $2.00 per hundred. Postage extra.

LANGUAGE, COMPOSITION AND GRAMMAR

I. Charters Diagnostic Language and Grammar Tests.

1. *Description.* These tests have been formed from material gathered in a study of oral errors in English made by the school children in Kansas City and Detroit. There are separate editions for (a) language and (b) language and grammar. The language tests are four in number, Pronouns, Verbs A, Verbs B, and Miscellaneous. There are three language and grammar tests, Pronouns, Verbs A and Miscellaneous. In the language tests the word forms are presented in sentences some of which are right and some of which are wrong. The children are instructed to supply the correct word forms. In the grammar tests they are asked to give the grammatical rule upon which the correct form is based.

2. *Function.* As the titles indicate, the purpose of these tests is mainly diagnostic. The teacher is able to ascertain the particular language errors, which the children are unable to correct or for which they cannot give the rule upon which the correction is based. The tests have also a general function.

3. *Time Required.* It has been found that it requires from 15 to 25 minutes for the children to complete one of the language tests. The time required in the scoring of the test is approximately equivalent to that required to mark 40 sentences in language. The language and grammar tests require more time.

4. The language tests may be used in grades 3 to 8 inclusive. The language and grammar tests are not suitable for use below the seventh grade.

5. One test is required for each pupil in the grades. There are two forms of each test. Indicate clearly the form and edition (language, or language and grammar) of the tests desired.

6. *References.* None available.

7. *Price*: Including complete directions and score sheets, 55 cents per hundred for language edition of any test; $1.15 per hundred for the language and grammar edition.

II. Nassau County Supplement to the Hillegas Scale.

1. *Description.* This scale is a simplification and improvement of the original Hillegas Composition Scale. It consists of a series of compositions written under controlled conditions and arranged in order of general merit.

2. *Function.* This scale was designed to yield a measure of the general merit of compositions written under standard conditions.

3. *Time Required.* Eighteen minutes are allowed for writing the compositions. The rating of the compositions by means of the scale requires no more time than is usually required for "grading" compositions. The length of time required can be reduced by practice. As experience increases the work will be done more accurately.

4. This scale may be used in grades 4 to 12 inclusive.

5. Only one scale is needed for each teacher. One class record sheet is needed for each grade. Give number of scales and class record sheets desired.

6. *References.* Hillegas, M. B.: "A Scale for the Measurement of Quality in English Composition by Young People," Teachers College Record, 13: 331-384, September, 1912; Trabue, M. R.: "Supplementing the Hillegas Scale," Teachers College Record, 19:51-85, January, 1917; Theissen, W. W.: "Improving Teachers' Estimates of Composition Specimens with the Aid of the Trabue Nassau County Scale," School and Society, 7: 143-150, February, 1918; Strayer, G. D.: "Some Problems in City School Administration," pp. 155-163, Yonkers-on-the-Hudson, New York, World Book Co., "The Composition Tests," Report of a Survey of the School System of St. Paul, Minnesota, 1917,

pp. 436-481; Parker F. E. and Courtis, S. A. "The Value of Measurements," English Journal, 8: 203-208, April 1919.

7. *Price*: Including complete directions and score sheets, 10 cents each, extra class record sheets, one cent each. Postage extra.

III. Starch Punctuation Scale. 1. *Description.* This scale consists of a series of groups of sentences which are to be punctuated. The groups are arranged in order of difficulty.

2. *Function.* This scale is designed to furnish a general measure of the pupil's ability to punctuate sentences.

3. *Time Required.* Accurate data are not available on this point. A class of children ought to complete the test in less than half an hour and the teacher should be able to score the papers with no more effort than is needed to mark the ordinary written language lesson of the same length.

4. The scale may be used in grades 5 to 8. It will also be useful in the high school.

5. One copy of the scale is needed for each pupil. There is only one form. Give the number of pupils to be tested.

6. *References.* Starch, D.: "The Measurement of Achievement in English Grammar," Journal of Educational Psychology, 6:615-626, December 1915.

7. *Price*: Including directions and score sheets, 70 cents per hundred. Postage extra.

IV. Trabue Language Scales. 1. *Description.* These scales are composed of completion sentences which are arranged in order of difficulty from the very simple to the very difficult. There are four scales for the elementary school, B, C, D, and E, and two scales for the high school, L and M.

2. *Function.* These tests aim to measure general language ability. These scales have often been used as intelligence tests. The scores of individual pupils on single scales are not highly reliable; hence several scales must be used in order to obtain a satisfactory measure of individuals.

3. *Time Required.* Each of the scales for the elementary grades requires seven minutes for its administration. Those for the high school take but five minutes each. The test papers may be scored at the rate of 50 an hour.

4. As has been implied above scales B, C, D, and E may be used in the elementary school and scales L and M in the high school.

5. The Bureau offers scales B, C, L, and M. One copy is required for each pupil. Give the number of elementary school pupils and the number of high school pupils to be tested. State which scales (B, C, L, or M) are desired.

6. *References.* Trabue, M. R.: "Completion-Test Language Scales," New York: Teachers College, Columbia University, 1916; Uhl, U. L.: "Mentality Tests for College Freshmen," Journal of Educational Psychology, 10:13-28, January, 1919.

7. *Price*: Including complete directions and class record sheets, 60 cents per hundred. Postage extra.

V. Willing Scale for Measuring Written Composition.

1. *Description.* This scale consists of eight compositions arranged in order of both "form value" and "story value." All of the compositions were written by school children under controlled conditions.

2. *Function.* The function of this scale differs from that of the Nassau County Scale in that it yields a measure of form (spelling, punctuation, and syntax) separate from the quality or "story value."

3. *Time Required.* Twenty-five minutes are required for writing the compositions. The rating of the compositions by means of the scale requires no more time than is usually required for "grading" them in the ordinary teaching routine. The length of time required can be reduced by practice. With added experience the work will be done more accurately.

4. This scale may be used in grades 4 to 12 inclusive.

5. Only one scale is needed for each teacher. One class record sheet is needed for each class. Give the number of scales and class record sheets desired.

6. *References.* Willing, M. H.: The Measurement of Written Composition in Grades 4 to 8," The English Journal, 7:193-252, March, 1918.

7. *Price*: Including complete directions and class score sheets, 5 cents each, extra class record sheets, one cent each. Postage extra.

READING

I. Courtis Silent Reading Test No. 2. 1. *Description.* This silent reading test consists of a simple and interesting connected story. The comprehension is measured by means of questions independently of the rate of reading.

2. *Function.* The function of this test is suggested by its description. It is to measure the rate and degree of comprehension resulting from the reading of a simple, interesting, connected story.

3. *Time Required.* Three minutes are allowed for the reading and five minutes for answering the questions. It is recommended by the author of the test that the scoring be done by the pupils, but it can be done by the teacher in two hours for a class of 25 pupils.

4. This test is to be used in grades 2 to 6 only.

5. There are three forms of the test. The Bureau carries Forms 1 and 3. In ordering specify which form is desired and give the number of pupils to be tested.

6. *References.* Forty-sixth Annual Report, Board of Education, Kansas City, Missouri, pp. 79-85, 1917.

7. *Price*: Including complete directions and class record sheets, $2.00 per hundred. Postage extra.

II. Gray Oral Reading Test. 1. *Description.* This is the most widely used oral reading test. It consists of a series of standardized paragraphs arranged in order of increasing difficulty.

2. *Function.* This test is designed to yield a general measure of oral reading ability, including both the rate of reading and the quality. However, the data of the test may be so tabulated as to secure a diagnosis of oral reading ability.

3. *Time Required.* The pupils must be tested individually in a room apart from those who have not been tested. The time required for a pupil varies, but generally is about five minutes. The test is scored by the examiner as the pupil reads. The scores of a class of 25 pupils can be tabulated in less than an hour.

4. The test can be used in grades 2 to 8.

5. Only one form is available. One copy of the test is needed for each pupil.

6. *References.* Gray, W. S.: "Studies of Elementary School Reading through Standardized Tests," Supplementary Educational Monograph, Vol. I, No. 1, University of Chicago Press, 1917; Gray, W. S.: "Reading," Survey of the St. Louis Public Schools, St. Louis, Board of Education, 1917; Gray, W. S.: "The Use of Tests in Improving Instruction," Elementary School Journal, 19:121-143, October, 1918.

7. *Price*: Including complete directions and score sheet 70 cents per hundred. Postage extra.

III. Gray Silent Reading Tests. 1. *Description.* These tests consist of three selections, "Tiny Tad" for grades 2 and 3, "The Grasshoppers" for grades 4, 5, and 6, and "Ancient Ships" for grades 7 and 8. The comprehension is measured both by reproduction and by answering questions. The author states: "For the purpose of this thesis it was desirable to devise a test which would be as accurate as possible. Wherever questions arose, ease of giving the tests was usually sacrificed for increased accuracy in results."

2. *Function.* These tests were designed to yield a measure of the rate and comprehension of the silent reading of connected material. The rate is measured independently from the comprehension. These tests are recommended for measuring the silent reading ability of backward pupils and other cases requiring individual attention.

3. *Time Required.* It requires about five minutes to give the test to each pupil. Scoring the reproductions and answers requires additional time.

4. The tests can be used in grades specified in the first paragraph.

5. Since the same test is not used in all grades it is necessary to give the number of pupils to be tested in each grade. Only one form of the tests is available.

6. *References.* Gray, W. S.: "Studies of Elementary School Reading through Standardized Tests," Supplementary Educational Monographs, Vol. I, No. 1, University of Chicago Press, 1917; Gray, W. S.: "Reading," Survey of the St. Louis Public Schools, St. Louis, Board of Education, 1917. Annual Report of the Board of Education, East Side, Aurora, Illinois, pp. 66-68, 1916.

7. *Price*: Any one of the three selections, two cents each. A reasonable provision will be one copy of a selection for each ten pupils tested. Reproduction sheets, (one for each pupil), 55 cents per hundred. Postage extra.

IV. Monroe Standardized Silent Reading Tests. 1. *Description.* These tests consist of a number of exercises, each of which requires the pupil to read a paragraph and answer a simple question upon it. The paragraphs were taken from school readers and other books children read. These tests have been widely used.

2. *Function.* These tests yield measures of the rate of reading and the degree of comprehension. Since the rate and comprehension are measured concurrently, the rate score is the rate of reading plus answering the questions. The material read is not connected in thought from one paragraph to the next. Thus, the rate score has a different meaning than the rate score in the Courtis Silent Reading Test N. 2, and Gray Silent Reading Tests.

3. *Time Required.* The test requires five minutes to give. The test papers of a class of 25 pupils can be scored and the scores tabulated within one hour or less.

4. Test I is used in grades 3, 4, and 5, Test II in grades 6, 7, and 8, and Test III in grades 9, 10, 11, and 12.

5. Since different tests are used in different grades, it is necessary to specify for what grades the tests are desired. There are three forms of each test except Test III which can be had in only two forms. Give the number of pupils to be tested in each grade and the form desired.

6. *References.* Monroe, Walter S.: "Monroe's Standardized Silent Reading Tests," Journal of Educational Psychology, 9: 303-312, June, 1918.

7. *Price*: Including complete directions and class record sheets, 60 cents per hundred. Postage extra.

V. Thorndike Scale Alpha 2 for Measuring the Understanding of Sentences. 1. *Description.* This scale consists of a series of paragraphs and questions about the paragraphs arranged in order of difficulty of comprehension.

2. *Function.* This scale measures only the degree of comprehension and does this in terms of the difficulty of the exer-

cises which the pupil is able to do 80 per cent correctly. No measure of the rate of reading is secured.

3. *Time Required.* Twenty minutes are required to give the test. The scoring of the test papers and the tabulation of the scores are more difficult than for most tests but detailed directions and tables for facilitating this work are furnished with the test.

4. Part I is used in grades 3 to 5 and Part II in grades 6 to 12.

5. Since the same exercises are not used for all grades it is necessary to specify the grades in which it is to be used. Only one form is available· Give the number of pupils to be tested.

6. *References.* Thorndike, E. L.: "An Improved Scale for Measuring Ability in Reading," Teachers College Record, 16: 445-467, 17: 40-67, November 1915, January 1916; Thorndike, E. L.: "Reading as Reasoning: A Study in Mistakes in Paragraph Reading," Journal of Educational Psychology, 8:323-332, June 1917; The Madison Public Schools, Sixty-second Annual Report, 1916-17, Madison, Wisconsin, pp. 56-59; Seventh Report of the Board of Education, Louisville, Kentucky, pp. 36-39, 1917-1918.

7. *Price*: Including complete directions and class record sheets, $1.00 per hundred for each part. Postage extra.

VI. Thorndike Visual Vocabulary Scales. 1. *Description.* Each of these scales consists of lists of words arranged in order of difficulty of meaning. The test requires the pupil to indicate the meaning of each word.

2. *Function.* These scales are designed to measure the pupil's knowledge of vocabulary which is an important factor in silent reading.

3. *Time Required.* It requires 20 minutes to give the test. The scoring requires about one hour for a class of ordinary size· The recording of the scores and calculation of the class score is facilitated by a record sheet and directions which are furnished with the test.

4. These scales may be used in grades 3 to 8.

5. There are four parallel forms, A2-x, A2-y, B-x, and B-y. In ordering state which form is desired. Give the number of pupils to be tested.

6. *References.* Thorndike, E. L.: "The Measurement of Achievement in Reading: Word Knowledge," Teachers College Record, 17:430-454, November 1916; "Report of a Survey of Public Education in Nassau County New York," University of the State of New York Bulletin, pp. 169-172, 1918.

7. *Price*: Including complete directions and class record sheets, 70 cents per hundred. Postage extra.

SPELLING

I. Ayres Spelling Scale. 1. *Description.* A thousand words constituting a fundamental English vocabulary have been grouped in this scale according to their spelling difficulty, understanding that spelling difficulty is indicated by the per cent of correct spelling. The words were selected from four important reports on the frequency with which words are used in letters, newspaper articles, the English Bible, and various authors. They, therefore, constitute a fundamental vocabulary. The approximate per cent of correct spelling to be expected in each grade is shown for each word.

2· *Function.* This scale is not a test but merely a classified list of words from which selections may be made for a standardized test. Since these words are those most commonly used, a test chosen from the scale measures the ability of the pupils to spell words which they probable use in their written language.

3. *Time Required.* The time required will depend upon the number of words used.

4. This scale may be used in grades 2 to 8.

5. Only one scale is needed for each teacher. One class record sheet is needed for each grade. In ordering give the number of scales and class record sheets desired.

6. *References.* Ayres, L. P.: "A Measuring Scale for Ability in Spelling," New York, Russell Sage Foundation, "Spelling Test," Research and Measurement Bulletin No. 1, Pittsburgh, 1918; "The Results of a Spelling Test," Cleveland, Board of Education, Division of Reference and Research, Bulletin No. 2, 1918.

7. *Price*: Including complete directions and one class record sheet, 7 cents each, extra class record sheets, one cent each. Postage extra.

II. Buckingham Extension of Ayres Spelling Scale. 1. *Description.* This extension adds 509 words to the original Ayres Spelling Scale making a scale of 1509 words. The additions are mainly at the more difficult end of the scale and will thus add to its value especially for the grammar grades and the high school. The difficulty of each of the new words has been determined and approximates the values for the different grades printed above the column in which the word is placed·

2. *Function.* This scale provides a classified list of words from which a teacher may select a spelling test of known difficulty.

3. *Time Required.* The time required will depend upon the number of words used.

4. This scale is planned for use in grades 2 to 9. Indeed, the additional words make its use practicable throughout the high school.

5. Only one scale is needed for each teacher. One class record sheet is needed for each grade. In ordering give the number of scales and class record sheets desired.

6. *References.* None available.

7. *Price*: Including complete directions and one class record sheet, 8 cents; extra class record sheets one cent each. Postage extra.

III. Monroe Timed Sentence Spelling Test. 1. *Description.* Each of these tests consists of 50 words taken form suitable columns of the Ayres Spelling Scale and embodied in sentences· The sentences are then so arranged that they may be dictated at approximately the normal rate of writing in each of the grades.

2. *Function.* The function of a timed sentence spelling test is to secure a more valid measure of a pupil's real spelling ability than is secured by means of words dictated in lists. Pupils normally spelling words as they occur in sentences have their attention fixed more upon what they are writing than upon the spelling. This condition is approximated by a timed sentence spelling test.

3. *Time Required.* The time varies with the diflerent grades and is only a little greater than would be used for a dictated list of 50 words.

4. Test I is used in grades 3 and 4, Test II in grades 5 and 6, and Test III in grades 7 and 8 and in the high school.

5. In ordering specify the grades in which the test is to be given unless the complete set is ordered. Only one copy is needed for each teacher. One class record sheet is needed for each grade. One form is now available. A second form will be devised when there is demand for it.

6· *References.* None available.

7. *Price*: Including three class record sheets, 8 cents per set, extra class record sheets one cent each. Postage extra.

CHAPTER VII. STANDARIZED TESTS FOR HIGH SCHOOL SUBJECTS

Certain series of tests described for the elementary school have a test designed for the high school. See Monroe's Standardized Silent Reading Tests and Thorndike's Scale Alpha 2 for Measuring the Understanding of Sentences. Other tests described for the elementary school can be used with profit in the high school. This is particularly true for the tests in composition, language, and grammar.

ALGEBRA

I. Hotz First-Year Algebra Scales. 1. *Description.* These scales consist of five lists of algebraic exercises: (1) addition and subtraction, (2) multiplication and division, (3) equation and formula, (4) problems, and (5) graphs. In each list the exercises are arranged in order of difficulty.

2. *Function.* These scales are designed to measure the ability of pupils in terms of the degree of difficulty of the exercises which just 50 per cent of the class are able to solve correctly. They are instruments for measuring the abilities of classes rather than of individual pupils. No measure of rate of work is secured.

3. *Time Required.* Pupils are allowed 110 minutes for the five tests of Series A and 200 minutes for Series B which is simply more elaborate than Series A. Additional time is required for scoring and tabulating. ;

4. These scales may be used as early as the end of the third month of the first year.

5. Only Series A is kept in stock by the Bureau of Educational Research. Series B can be obtained from the Bureau of Publications, Teachers College, New York City. In ordering give the number of pupils to be tested.

6. *References.* Hotz, H. G.: "First-Year Algebra Scales," New York, Teachers College, Columbia University, 1918.

7. *Price*: Including complete directions and class record sheets, Tests 1, 2, 3, and 4, 70 cents per hundred each test. Test 5, $1.25 per hundred. Postage extra.

GEOMETRY

I. Minnick Geometry Tests. 1. *Description.* There are five tests. Test A, drawing figures for propositions; Test B, stating what is given and what is to be proved in propositions; Test C, stating facts about figures given; Test D, organizing proof from given data.

2. *Function.* Tests A, B, C, and D are diagnostic.

3. *Time Required.* A maximum allowance of 30 minutes each is given to each test. The scoring requires about the same time as do ordinary geometry examination papers.

4. These tests may be used with all classes that have studied at least two books of plane geometry.

5. One copy of each test is required for each pupil. Only one form is offered.

6. *References.* Minnick, J. H.: "Minnick's Geometry Scale for Measuring Pupil's Ability to Demonstrate Geometrical Theorems," School Review 27:105-109, February 1919; Minnick, J. H.: "An Investigation of Certain Abilities Fundamental to the Study of Geometry," (A thesis) Philadelphia, University of Pennsylvania, 1918.

7. *Price*: Including complete directions and class record sheets, $2.50 per hundred for each test. Postage extra.

HISTORY

I. Sackett Scale in Ancient History. 1. *Description.* This scale is composed of six tests based on carefully selected information in ancient history.

2. *Function.* The scale aims to give a general measure of the pupil's grasp of the facts of ancient history.

3. *Time Required.* The pupils are given exactly 29 minutes to answer the questions in the six tests. The scoring takes about the time required to mark the papers in a history examination of six questions.

4. This test may be given to pupils who have completed the study of a year of ancient history.

5. One copy of the test is needed for each pupil in the class.

6. *References.* Sackett, L. W.: "A Scale in Ancient History," Journal of Educational Psychology, 8:285-293, May 1917.

7. *Price*: Including class record sheet, $2.00 per hundred. Postage extra.

LATIN

I. Henmon Latin Tests. 1. *Description.* These tests are composed of two types of exercises which have been graded in order of difficulty. One set of exercises furnishes a measure of Latin vocabulary. The other set gives a measure of ability to translate Latin sentences. There are five vocabulary lists, A, B, C, D, and X and three sentence tests I, II, and III. Only tests D, X, II, and III are carried in stock.

2. *Function.* These tests are diagnostic in that they discriminate between a pupil's vocabulary and his ability to translate sentences. They are general in each of these fields.

3. *Time Required.* No data are available on this point.

4. The vocabulary test may be used with pupils who have had first-year Latin. Sentence Tests I and II may also be used with pupils who have had a single year's study of the language. Test III should not be given until Latin has been studied three or four years.

5. One copy of each test is needed for each pupil in the class. Indicate which tests are desired and give the number of pupils to be tested.

6. *References.* Henmon, V. A. C.: "The Measurement of Ability in Latin," Journal of Educational Psychology, 8:515-589, November and December, 1917.

7. Price: Including complete directions and score sheets, $1.00 per hundred each test. Postage extra.

PHYSICS

I. Starch Physics Test. 1. *Description.* This test is composed of 75 sentences which are arranged in the form of completion tests. The degree to which the pupils have mastered the work in physics covered by these sentences is shown by their ability to fill in the missing words, phrases, and numbers. The test is broken up into topics and standard scores for each topic are printed on the test sheet. There are 28 sentences on mechanics, 8 on heat, 9 on sound, 9 on light, and 21 on magnetism and electricity.

2. *Function.* The test is designed to give a general measure of the pupil's mastery of the different topics in physics.

3. *Time Required.* The administration and the scoring of the test are comparatively simple. Neither requires as much time as would be given to the ordinary physics examination.

4. The test may be used with pupils who have completed the study of physics, or it may be given in parts as each topic is completed.

5. One test is required for each pupil. Only one form is available. Give the number of pupils to be tested.

6. *References.* None are available.

7. *Price*: Including complete directions and class record sheets, $3.75 per hundred. Postage extra.

CHAPTER VIII. MENTAL TESTS.

There are a number of group intelligence scales and tests on the market. None of them, however, is sufficiently reliable so that it can be used alone to secure accurate measures of the intelligence of individual children. To be safe, any person who administers group tests or scales should not be content with one measure, but should secure as many as his resources permit. In all probability the use of two or three of the better group intelligence tests will give a fairly reliable index of a child's intelligence, although the results of a single test might do him appreciable injustice.

The results of the use of these scales will assist in properly classifying children in the public schools. In this work several factors should be taken into consideration in addition to the crude scores. The mental test scores should be converted into mental ages and the mental ages should be considered in relation to the chronological ages. The simplest way to do this is to divide the mental age by the chronological age, thus securing a so-called intelligence quotient. If the children are more than 16 years of age (chronologically), 16 should be used as the divisor, since it has been assumed that mental maturity is reached at approximately the age of 16. The best use to make of these facts, then, is to prepare a table showing the relation between intelligence quotient and age for the grade. All children who are retarded—i. e., older chronologically than the normal age for the grade—should be canvassed to ascertain if they are qualified for extra promotions. If a child who shows an approximate normal intelligence quotient (1.00) is a year or two older than the average of his grade, and if he shows a reasonably satisfactory record in scholarship, he should be promoted to a higher grade and given such extra attention as is needed to bridge the gap caused by the extra promotion.

Mental tests may be used also to classify the children into homogeneous learning groups. The bright children and the dull children may each be put into sections where they will receive special treatment. The separation of the bright and the dull from the average has been found to facilitate the instruction of all three groups. The argument which school people sometimes advance, namely that the presences of bright pupils gives tone to a class, or in other words, that the average and the slow profit by the contact with better minds, is mainly an argument of conservatives opposed to innovation. The classification of children according to their mental ages is made possible by the tentative standards which are furnished with each of the tests and scales offered for sale by the Bureau.

Those who administer mental tests will frequently meet situations where the test that is being used with a group is too difficult for some of the children. These pupils who make zero scores on some of the individual tests in the scales or will make zero scores on the entire scale. Wherever this happens the thing to do is to give the individual in question a chance to take a scale designed for more elementary mental types. There are, for example, children in the high school who are so mediocre in intelligence that their scores are much below the averages for children in the seventh and eighth grades. If high school pupils show up very poorly when tested with scales designed for the high school, they should be given scales designed for the seventh and eighth grades. Similarly, many children will be found in the fourth and fifth grades who can do nothing with the literary material prepared for use in these grades. Consequently, they should be tested with material designed for primary grades. Children whose mentality is too low to be measured by scales for the primary grades are probably so markedly inferior that they cannot profit by the ordinary work of the public schools. It is possible that they might do some of the kindergarten tasks, but they are hardly mature enough mentally to undertake the work of the primary grades.

I. Holley Picture Completion Test for Primary Grades.

1· *Description.* This test is composed of pictures from which parts were omitted when they were drawn. The children are expected to complete the pictures by adding the missing parts. The pictures used in this scales have been evaluated. They are arranged approximately in order of difficulty from the simplest to the most difficult.

2. *Time Required.* The children are given five minutes in which to complete the pictures. Approximately ten minutes will be required by teachers in the second and third grades for the distribution of the papers, the filling in of the blanks, and the explanation of the directions which the children are given. The papers can be scored at the rate of 75 to 100 per hour.

3. Only one form of this test is available. It is suitable for use in the first three grades.

4. One copy of the test is needed for each pupil. In ordering give the number of pupils to be tested.

.5 *References.* None available at this time.

6. *Price*: Including complete directions and score sheets, $1·25 per hundred. Postage extra.

II. Holley Sentence Vocabulary Scale. 1. *Description.* This test has been devised with the Terman Stanford-Binet Vocabulary as a basis. The words of this well-known vocabulary are put in sentences the last word of which is found among four words to the right. The children indicate the correct responses by underlining the word which completes the sentence. This scale has been found to be a fair measure of intelligence.

2. *Time Required.* These scales are given without time limits and the time required varies from about 35 minutes for the third grade to about 20 minutes for the twelfth grade. The

papers can be scored at the rate of approximately 35 an hour.

3. The vocabulary scale is prepared in two series: Series 3A for grades 3 to 8, and Series 3B for grades 7 to 12.

4. One copy of the test is needed for each pupil to be tested. In ordering give the series and the number of children to be tested.

5. *References.* None available.

6. *Price*: Including complete directions and score sheets, 75 cents per hnudred. Postage extra.

III. Pressey Primer Scale. 1. *Description.* This scale is composed of four tests a dot pattern, a classification, a form-board, and an absurdities test·

2. *Time Required.* Approximately 25 minutes are required to administer it to a group of children and the papers can be scored at the rate of 35 per hour.

3. The scale is intended for use in the first three grades. Only one form is available at present.

4. One copy of the scale is needed for each pupil to be tested.

5. *References.* None available.

6. *Price*: Including complete directions and score sheets, $1.00 per hundred.

MENTAL TESTS NOT SOLD BY THE BUREAU

I. Otis Group Intelligence Scale. 1. *Description.* The scale is composed of ten tests which measure various phases of mental ability.

2. *Time Required.* Approximately 70 minutes are required to administer it to a group of children. The papers can be scored at about the rate of 13 an hour.

3· Two forms of the scale are available. They may be used in grades 6 to 12. The author suggests their use in lower grades but only the brighter children can cope with them successfully below the sixth grade.

4. One copy of the scale is needed for each pupil to be tested. In ordering indicate the form desired and give the number of sets of 25 scales needed.

5. *References.* Otis, Arthur S.: "An Absolute Point Scale for the Group Measurement of Intelligence," Journal of Educational Psychology, pp. 239-261, 333-348, 1918.

6. *Publisher.* This scale is sold by the World Book Company, Yonkers-on-the-Hudson, N. Y.

II. Virginia Delta 1. 1. *Description.* This scale was devised by the Virginia Education Commission and was used in the mental survey of the schools in Virginia. It consists of six parts, composed of different types of materials.

2. *Time Required.* Approximately 30 minutes are required to administer this scale to a group of children. These tests can be scored at the rate of about 20 per hour·

3. Only one form of the scale is available at present.

4. One copy of the scale is required for each pupil taking the test. When ordering give the number of pupils to be tested.

5. *References.* None available.

6. *Publisher.* Published by the World Book Company, Yonkers-on-the-Hudson, N. Y.

III. Virginia Delta 7. 1. *Description.* This is another scale devised by the Virginia Education Commission. It was intended for use in the primary grades. It is composed of 12 tests. It is a new scale which promises to be of value for the primary grades.

2. *Publisher.* Published by the World Book Company, Yonkers-on-the-Hudson, N. Y.

PRICE LISTS OF EDUCATIONAL TESTS, JANUARY, 1920

The Bureau of Educational Research is prepared to furnish the test materials listed on the pages of this folder. The following statements should be read carefully before making an order.

How to order standardized tests. In ordering a standardized test of which one copy is needed for each pupil, give the number of pupils in each grade which are to be tested. This is *important* because in the case of certain tests, the exercises vary for different grades. Unless one is familiar with the details of the test or scale being ordered, it will be well to refer to the description of the test, particularly to item 5 in each description. Where more than one form is available, specify the form desired One original and one duplicate record sheet will be furnished with every 25 test papers ordered. A limited number of additional copies will be supplied upon request. Class record sheets for one grade will be furnished with each scale. Where scales are to be used for more than one grade or by several teachers additional copies may be ordered. Price for additional copies will be furnished upon request. Cash should not accompany the order unless the total is for less than 25 cents. In this case remittance should be sent with the order and stamps will be received. A statement will be sent from the University Business Office for all amounts over 25 cents. Address all orders and correspondence, however, to the Bureau of Educational Research, University of Illinois, Urbana, Illinois.

Prices. All prices must be considered tentative. Industrial conditions are so unsettled that it is impossible to quote permanent prices at this time. Tests will be sold at cost and prices will be readjusted as conditions change. In many instances tests which originate outside of the Bureau are sold at a slight advance over the prices quoted by those who publish them. This advance has been made necessary because the Bureau has added accessory material to enable teachers and superintendents to give the tests intelligently. In no cases do the prices for tests in quantity include transportation charges.

All materials listed by the hundred will be charged for at the sample rate if less than 15 of each kind are ordered; but orders for different parts of the same test will be received as orders for only one kind. Thus, a person will receive quantity prices on less than 15 copies of Monroe's Standardized Silent Reading Test for grades 3 to 5, if at the same time he orders a sufficient number of copies for grades 6 to 8, or 9 to 12 or both to make a total of more than 15 copies.

Accessory materials. All accessory material will be supplied with every order unless the person ordering makes a specific statement that they are not wanted. Prices of tests without accessory material will be submitted upon request.

Sample test materials. The Bureau of Educational Research is not prepared to furnish free samples of tests listed in this folder. Sample copies of any of the tests including the accompanying accessories will be sent postpaid for the indicated price. Attention is called to the sample packages which have been assembled to meet the demands of superintendents and teachers.

For the convenience of persons unfamiliar with tests, the Bureau has assembled six sample packages which contain assortments of the tests so arranged as to meet different needs. Each test is accompanied by all existing accessory material.

Sample Package A contains one test fo reach of the subjects in the elementary school. The test included are the ones believed to be best suited for teachers who are just beginning to make use of educational tests.

Courtis Standard Research Tests in Arithmetic, Series B
Monroe Standardized Reasoning Tests in Arithmetic
Ayres Handwriting Scale, "Gettysburg Edition"
Buckingham Extension of the Ayres Spelling Scale
Courtis Supervisory Test in Geography
Harlan Test of Information in American History
Charters Diagnostic Language and Grammar Test, Pronouns, Verbs A, Verbs B, and Miscellaneous
Willing Scale for Measuring Written Composition
Gray Oral Reading Test
Monroe Standardized Silent Reading Tests I and II
Thorndike Visual Vocabulary Scale
Price: 60 cents postpaid

Sample Package B is no longer sold.

Sample Package C is composed of material for use in the high school. It consists of the following:

Hotz First-year Algebra Scale
Minnick Geometry Tests
Sackett Scale in United States History
Sackett Scale in Ancient History
Henmon Latin Tests:
  a) Vocabulary Test (D)
  b) Vocabulary Test (X)
  c) Sentence Test II
  d) Sentence Test III
Willing Scale for Measuring Written Compostion
Nassau County Supplement to the Hillegas Scale
Thorndike Scale Alpha II for Measuring the Understanding of Sentences
Starch Physics Test
Monroe Silent Reading Test No. III
Jones Union Science Tests—Physics—(1 set)
Price: $1:00 postpaid.

Sample Package D contains tests in arithmetic, reading ,and spelling. It consists of the following:

Arithmetic

Buckingham Scale for Problems in Arithmetic, Divisions 1, 2, and 3.
Cleveland Survey Arithmetic Tests
Courtis Standard Research Tests in Arithmetic, Series B
Courtis Supervisory Tests in Arithmetic, Tests A and B
Monroe Diagnostic Tests in Arithmetic, Parts 1, 2, 3, and 4
Monroe Standardized Reasoning Tests in Arithmetic, Tests I, II, and III
Woody Arithmetic Scales, Series B, four tests

Reading

Courtis Silent Reading Test, No. 2
Gray Oral Reading Test
Gray Silent Reading Tests
Monroe Standardized Silent Reading Tests I and II
Thorndike Scale Alpha II for Measuring the Understanding of Sentences, Parts 1 and 2
Thorndike Vocabulary Scales

Spelling

Ayres Spelling Scale
Buckingham Extension of the Ayres Spelling Scale
Monroe Timed Sentence Spelling Tests

Price: $1:20 postpaid

Sample Package E contains tests in geography, United States History, language, and grammar. It consists of the following:

Geography

Courtis Supervisory Test in Geography
Hahn-Lackey Geography Scale

History

Harlan Test of Information in American History
Sackett Scale in United States History

Language and Grammar

Charters Diagnostic Language and Grammar Tests, Pronouns and Verbs A, Verbs B, and Miscellaneous.
Starch Punctuation Scale

Price: 65 cents postpaid

Sample Package F contains tests in handwriting and English composition. It consists of the following:

English Composition

Nassau County Supplement to the Hillegas Scale
Willing Scale for Measuring Written Composition

Handwriting

Ayres Handwriting Scale "Gettysburg Edition"
Freeman Chart for Diagnosing Faults in Handwriting
Gray Standard Score Card for Measuring Handwriting, Forms I, II, III, and IV
Thorndike Handwriting Scale

Price: 70 cents postpaid

Sample Package G is composed of intelligence tests. It consists of the following:

Holley Picture Completion Test for Primary Grades
Holley Sentence Vocabulary Scale
Pressey Primer Scale

Pressey Cross-Out Test
Whipple Group Test for Grammar Grades
Otis Group Intelligence Tests

Price: 70 cents postpaid

When they can be obtained there will be added to this package the Theisen-Flemming Classification Test and the Virginia Group Intelligence Scale, Delta 1. The price of the package will be increased to cover the expense of these additional tests.

ORDER BLANK AND SUMMARY OF PRICES

The prices given below do not include transportation charges except in the case of sample copies and sample packages. A sample set of any test includes one copy of each test or division of the series, together with one copy of each of the accessories. In case there is more than one form of the test, only one form is included in the sample test.

ELEMENTARY SCHOOL TESTS

Arithmetic

No. Copies

1. Buckingham Scale for Problems in Arithmetic

Sample Set, 7 cents
Division I, for grades 3 and 4, 70 cents per 100
Division II, for grades 5 and 6, 70 cents per 100
Division III, for grades 7 and 8, 70 cents per 100

2. Cleveland Survey Arithmetic Tests, for grades 3 to 8, $1.90 per 100

Sample Set, 9 cents

3. Courtis Standard Research Tests in Arithmetic, Series B

Sample Set, 22 cents
Form 1 for grades 4 to 8, $1.00 per 100
Form 3 for grades 4 to 8, $1.00 per 100

4. Monroe Diagnostic Tests in Arithmetic

Sample Set, 10 cents
Part I, Integers, for grades 4 to 8, 70 cents per 100
Part II, Integers, for grades 5 to 8, 70 cents per 100
Part III, Common Fractions, for grades 6 to 8, 70 cents per 100
Part IV, Decimal Fractions, for grades 6 to 8, 70 cents per 100

5. Monroe Standardized Reasoning Tests in Arithmetic

Form 1
Sample Set, 7 cents
Test I, for grades 4 and 5, 65 cents per 100
Test II, for grades 6 and 7, 65 cents per 100
Test III, for grade 8, 65 cents per 100
Form 2
Sample Set, 7 cents
Test I, for grades 4 and 5, 65 cents per 100
Test II, for grades 6 and 7, 65 cents per 100
Test III, for grade 8, 65 cents per 100

6. Woody Arithmetic Scales, Series B,

Sample Set, 6 cents
Four tests, 50 cents per 100 each test

Geography

1. Courtis Standard Supervisory Test in Geography, $1.00 per 100

Sample Set, 25 cents

	No. Copies
2. **Hahn-Lackey Geography Scale, 10 cents each**	
Extra record sheets, 1 cent each	

Handwriting

1. **Ayres Handwriting Scale "Gettysburg Edition," 7 cents each**	
Extra Record sheets, 1 cent each	
2. **Freeman Chart for Diagnosing Faults in Handwriting, 30 cents each**	
3. **Gray Standard Score Card for Measuring Handwriting**	
Sample Set, 15 cents	
Form I, 50 cents per 100	
Form II, 50 cents per 100	
Form III, 10 cents each	
Form IV, 30 cents per 100	
4. **Thorndike Handwriting Scale, 10 cents each**	
Extra record sheets, 1 cent each	

History

1. **Harlan Test for Information in American History, 75 cents per 100**	
Sample Set, 4 cents	
2. **Sackett Scale in United States History, $2.00 per 100**	
Sample Set, 6 cents	

Language, Grammar, and Composition.

1. **Charters Diagnostic Language Tests for grades 3 to 8, Form 1**	
Sample Set, 8 cents	
Pronouns, 55 cents per 100	
Verbs A, 55 cents per 100	
Verbs B, 55 cents per 100	
Miscellaneous, 55 cents per 100	
2. **Charters Diagnostic Language and Grammar Tests for grades 7 and 8, Form 1**	
Sample Set, 10 cents	
Pronouns, $1.15 per 100	
Verbs A, $1.15 per 100	
Miscellaneous, $1.15 per 100	
3. **Nassau County Supplement to the Hillegas Scale, 10 cents each**	
Extra class record sheets, 1 cent each	
4. **Starch Punctuation Scale, 70 cents per 100**	
Sample Set, 4 cents	
5. **Trabue Language Scales**	
Sample Set, 8 cents	
Scale B, 60 cents per 100	
Scale C, 60 cents per 100	
Scale L, 60 cents per 100	
Scale M, 60 cents per 100	

	No. Copies
6. Willing Scale for Measuring Written Composition, 5 cents each	
Extra record sheets, 1 cent each	

Reading

	No. Copies
1. Courtis Silent Reading Test No. 2	
Sample Set, 17 cents	
Form 1, for grades 2 to 6, $2.00 per 100	
Form 3, for grades 2 to 6, $2.00 per 100	
2. Gray Oral Reading, 70 cents per 100	
Sample Set, 4 cents	
3. Gray Silent Reading Tests	
Sample Set, 10 cents	
Selections	
"Tiny Tad" for grades 2 and 3, 2 cents each	
"Grasshoppers" for grades 4, 5, and 6, 2 cents each	
"Ancient Ships" for grades 7 and 8, 2 cents each	
Reproduction sheets	
"Tiny Tad" for grades 2 and 3, 55 cents per 100	
"Grasshoppers" for grades 4, 5, and 6, 55 cents per 100	
"Ancient Ships" for grades 7 and 8, 55 cents per 100	
4. Monroe Standardized Silent Reading Tests	
Sample Set, 6 cents	
Form 1	
Test I, for grades 3, 4, and 5, 60 cents per 100	
Test II, for grades 6, 7, and 8, 60 cents per 100	
Form 2	
Test I, for grades 3, 4, and 5, 60 cents per 100	
Test II, for grades 6, 7, and 8, 60 cents per 100	
Form 3	
Test I, for grades 3, 4, and 5, 60 cents per 100	
Test II, for grades 6, 7, and 8, 60 cents per 100	
5. Thorndike Scale Alpha 2 for Measuring the Understanding of Sentences	
Sample Set, 6 cents	
Part I, for grades 3 to 5, $1.00 per 100	
Part II, for grades 6 to 12, $1.00 per 100	
6. Thorndike Visual Vocabulary Scales	
Sample Set, 6 cents	
Form A2-x, 70 cents per 100	
Form A2-y, 70 cents per 100	
Form B-x, 70 cents per 100	
Form B-y, 70 cents per 100	

Spelling

	No. Copies
1. Ayres Spelling Scale, 7 cents each	
Extra record sheets, 1 cent each	
2. Buckingham Extension of the Ayres Spelling Scale, 8 cents each	
Extra record sheets, 1 cent each	
3. Monroe Timed Sentence Spelling Test	
Sample Set, 12 cents	
Five or more sets, 8 cents per set	
Extra record sheets, 1 cent each	

HIGH SCHOOL TESTS

In addition to tests specifically designed for use in the high school, a few tests for elementary school use are listed because it is believed they may be used profitably.

	No. Copies
Henmon Latin Tests	
Sample Set, 10 cents	
Vocabulary D, $1.00 per 100	
Vocabulary X, $1.00 per 100	
Sentence II, $1.00 per 100	
Sentence III, $1.00 per 100	
Hotz First-Year Algebra Scale	
Sample Set, 15 cents	
Test 1, 70 cents per 100	
Test 2, 70 cents per 100	
Test 3, 70 cents per 100	
Test 4, 70 cents per 100	
Test 5, 90 cents per 100	
Minnick Geometry Tests	
Sample Set, 15 cents	
Test A, $2.50 per 100	
Test B, $2.50 per 100	
Test C, $2.50 per 100	
Test D, $2.50 per 100	
Monroe Standardized Silent Reading	
Sample Set, 5 cents	
Form 1, Test III, 60 cents per 100	
Form 2, Test III, 60 cents per 100	
Nassau County Supplement to the Hillegas Scale, 10 cents each	
Extra record sheets, 1 cent each	
Sackett Scale in Ancient History, $2.00 per 100	
Sample Set, 6 cents	
Sackett Scale in United States History, $2.00 per 100	
Sample Set, 6 cents	
Starch Physics Test, $3.75 per 100	
Sample Set, 5 cents	
Thorndike Scale Alpha 2 for Measuring the Understanding of Sentences	
Sample Set, 5 cents	
Part II, $1.00 per 100	
Willing Scale for Measuring Written Composition, 5 cents each	
Extra record sheets, 1 cent each	
MENTAL TESTS	
Holley Picture Completion Test for Primary Grades, $1.25 per 100	
Sample Set, 5 cents	
Holley Sentence Vocabulary Scale	
Series 3A for grades 3 to 8, 75 cents per 100	
Series 3B for grades 7 to 12, 75 cents per 100	
Pressey Primer Scale for grades 1 to 3, $1.00 per 100	
Theisen-Flemming Classification Test, price to be announced	
Sample Set, price to be announced	
SAMPLE PACKAGES	
Sample Package A, 60 cents	
Sample Package C, $1.00	
Sample Package D, $1.20	
Sample Package E, 65 cents	
Sample Package F, 70 cents	
Sample Package G, 70 cents	

Address all orders and correspondence to the Bureau of Educational Research, University of Illinois, Urbana, Illinois.

BULLETINS OF THE BUREAU OF EDUCATIONAL RESEARCH

B. R. Buckingham, *Editor*

BUREAU OF EDUCATIONAL RESEARCH—BULLETIN NO. 3

STANDARD REQUIREMENTS FOR MEMORIZING LITERARY MATERIAL

BY

VELDA C. BAMESBERGER, A.M.

Research Assistant
Bureau of Educational Research
University of Illinois

Price 50 Cents

PUBLISHED BY THE UNIVERSITY OF ILLINOIS
URBANA

Copyright, 1920
By THE UNIVERSITY OF ILLINOIS

FOREWORD

This bulletin grew out of an investigation which has been carried on in the Bureau of Educational Research during the past two years. A part of the study afforded material for a Master's thesis which was submitted to the Graduate School of the University of Illinois in June, 1919. Sincere thanks are extended to Professor B. R. Buckingham, Director of the Bureau of Educational Research, under whose direction and supervision the investigation was made.

EDITORIAL INTRODUCTION

If we accept Arnold's definition of culture as "acquainting ourselves with the best that has been known and said in the world," we shall make no small effort to assure ourselves that that with which we become acquainted shall indeed be the best. If we transfer this idea to the school and seek to offer to children the best that has been said and done, we shall be very careful in our selection of teaching materials. The best thinking and doing of the race are recorded in books. A part of the record is so well made—so fine in thought and so beautiful in form—that we desire school children to acquire it in the exact words in which it is expressed.

It is clear that the body of such material cannot be large; and it is equally clear that it must be of unusual excellence. From the vast range of what we call literature we select as much of the best as we can for reading purposes. But far more rigorously selected material is chosen to be memorized—to be made a permanent possession.

The report presented in the following pages attempts to show of what this highly restricted body of material actually consists according to the practice of the schools from which information may be obtained. Opinions will differ as to the extent to which these schools have succeeded in selecting the best. The value of the poems, proverbs, and "memory gems" is not a primary consideration in the report. It is evident, however, that any desirable reform must be based upon present practice as a point of departure. If memory requirements are ill-advised, we must know what these requirements are in order to supplant them. If they are well suited to the purpose, we must know what they are in order to preserve them. It is, therefore, believed that to show the facts, and to present them in a useful way, will not only be of immediate practical value to teachers and school officers, but will also provide an indispensable basis upon which a more worthy selection of memory materials may subsequently be made.

B. R. BUCKINGHAM,

Director, Bureau of Educational Research,
University of Illinois.

TABLE OF CONTENTS

PART I—POETRY

PAGE

I The Problem 9

II Procedure 10

III Distribution of Poems 11

IV The Selected List 12

V Grading the Selected List 22

VI Verifying the Grading of the Selected List . . 39

VII Preferred Lists of Poems for Memory Work . . 43

VIII The Finding List 47

IX Authors Mentioned In The Selected List . . . 66

X Quantitative Standards 74

XI "Memory Gems" 79

PART II—PROSE

I Proverbs 80

II Biblical Material 83

III Longer Prose Selections 86

PART III—BIBLIOGRAPHIES

I Courses of Study 87

II Collections of Poetry 90

III Collections of Quotations 91

IV Readers 92

LIST OF TABLES

PAGE

I Distribution of Poems According to Frequency of Mention 11

II A Selected List of Poems and Parts of Poems Mentioned Five or More Times in Fifty Courses of Study 13

III A Suggested Grading of Peoms Mentioned Five or More Times 24

IV Lists of Poems for Memorizing in Grade I . . 31

V Lists of Poems for Memorizing in Grade II . . 32

VI Lists of Poems for Memorizing in Grade III . . 33

VII Lists of Poems for Memorizing in Grade IV . . 34

VIII Lists of Poems for Memorizing in Grade V . . 35

IX Lists of Poems for Memorizing in Grade VI . . 35

X Lists of Poems for Memorizing in Grade VII . . 36

XI Lists of Poems for Memorizing in Grade VIII . . 37

XII Correspondence in the Grading of the 111 Poems Common to the Selected List and Atherton's List 40

XIII Correspondence in the Grading of the 272 Poems Common to the Selected List and to the Elementary Readers 42

XIV Preferred Lists of Poems for Memory Work . . 43

XV The Selected List Organized as a Finding List . . 48

XVI Number of Poems by Authors and by Suggested Grades. Selected List 66

XVII Frequency of Mention of Each Author Represented in the Selected List 71

XVIII Minimum Quantitative Requirements for Memory Work 76

XIX Distribution of the Courses of Study According to the Number of Poems Required to be Memorized in Each Grade 79

STANDARD REQUIREMENTS FOR MEMORIZING LITERARY MATERIAL

PART I—POETRY

I—The Problem

The amount and character of the memory work required in English Literature in the elementary school came up as a practical problem in the Bureau of Educational Research. The question emanated from a Superintendent of Schools who wished to know standards pertaining to memory work. An attempt was made to find investigations bearing on the subject, and one study was found. In the January, 1914, number of the Elementary School Teacher, Lewis Atherton published an article entitled "Literary Selections Most Frequently Memorized in The Public School." His data were collected from thirty-four city and state courses of study. He listed the selections mentioned by three or more courses according to the grades in which they were most often placed.

The article, however, was no more than a good beginning. It did not contain quantitative standards for memory work—i.e., no answer to the question "How much memorizing of literary material is being required." It made no distinction between poetry and prose. It grouped but a relatively small part of the selections (those mentioned 10 or more times) into grade lists and omitted to provide a bibliography showing where the selections might be found in printed form. Complete data as to the frequency of mention of a given poem in each grade were not given, nor an arrangement of the poems in the order of their popularity. Moreover the study was carried out upon a relatively small number of courses of study, and had been finished at least five years before the problem came up for consideration by the Bureau of Educational Research. The problem was therefore thought worthy of a more extensive treatment than had been attempted in Atherton's article.

II—PROCEDURE

Two hundred or more courses of study were examined and about one hundred which contained references to memory work were chosen. Later the state courses of study were eliminated as well as some city courses whose statements regarding memory work lacked precision. The number was thus reduced to fifty of the better city courses of study, and it is upon these that this report is directly based. These were the courses which showed the most complete lists of memory material and which contained specific requirements as to the amount of memory work demanded. An attempt was made to include courses of study from cities in various parts of the country. Five were from cities having a population of more than 250,000, eleven from cities between 100,000 and 250,000, and thirty-four from cities between 25,000 and 100,000. The cities were distributed geographically as follows: twenty from the North Atlantic states, three from the South Atlantic states, eleven from the North Central states, eight from the South Central states, and eight from the Western states.*

The references to literary material other than poetry were so indefinite that only poems were studied in detail. Some courses of study mentioned prose selections. In most of these cases it was impossible to determine agreement. In Part II, however, an attempt has been made to show in a general way the common practice with respect to material other than poetry.

The nature of the requirements concerning memory work varied greatly among the different courses of study. Poems were indicated "for study" by some courses. Others gave lists of poems, a few of which were "required for memory," the rest being merely "suggested for memory." Still other courses gave very long lists from which the teacher might choose a specified number for memory work. A few courses required all the poems mentioned to be memorized. A record of the character of the mention accorded each selection was kept. It was thought that the information might be used in determining its "importance" for memory work. The grades in which poems were mentioned were also recorded in order that the poems might be arranged in graded lists. All the poems were located in collections of poetry and the first lines were copied.

---

*A complete list of the courses of study is shown in Part III of this report.

III—DISTRIBUTION OF POEMS

In all 2,435 different poems were mentioned in the fifty courses of study. Table I shows the distribution of poems according to frequency of mention. One thousand five hundred forty poems or about five-eights of the 2,435 poems were listed but once. A great many of these poems are anonymous, and some are very long. This latter fact may explain why most cities do not include them in their lists of memory selections. Sectional preferences are also evident to a certain extent. For example, a large number of Lanier's poems were mentioned by one southern city, while only one of them was popular enough to be mentioned by as many as five cities. Three hundred twenty-nine poems were mentioned five or more times. One hundred seventy-five poems were mentioned ten or more times; seventy-seven, twenty or more times; forty-six, twenty-five or more times; and twenty-six, thirty or more times.

TABLE I—DISTRIBUTION OF POEMS ACCORDING TO FREQUENCY OF MENTION

Times Mentioned	NUMBER OF POEMS	
	Mentioned the indicated number of times	Accumulated from the beginning
Over 45	1*	1
40—44	6	7
35—39	7	14
30—34	12	26
25—29	20	46
20—24	31	77
15—19	35	112
10—14	63	175
5—9	154	329
(9)	(18)	(193)
(8)	(25)	(218)
(7)	(30)	(248)
(6)	(31)	(279)
(5)	(50)	(329)
1—4	2106	2435
(4)	(96)	(425)
(3)	(155)	(588)
(2)	(315)	(895)
(1)	(1540)	(2435)
Total	2435	

*59

IV—THE SELECTED LIST

The 329 poems mentioned five or more times in the fifty courses of study constitute the so-called Selected List which forms the basis of this report. Table II contains this list arranged in the order of decreasing frequency of mention. The poem which was named the greatest number of times (59) was "America."

It sometimes happened that a given course of study mentioned the same poem in two or more grades. This was particularly true in the case of patriotic poems. In Table II the column headed "Cities Mentioning" indicates the number of different cities whose courses of study listed each poem. As will be observed, "America," although mentioned 59 times, was listed by but 30 of the 50 cities. Fifteen other poems were listed by a greater number of cities. It is rather remarkable that 20 out of 50 representative cities should have failed to mention "America," when listing poems which they desire their children to learn. Perhaps most if not all of these cities, however, required the memorizing of this poem in connection with the music work. The same condition holds true for the "Star Spangled Banner" (No. 8) which was mentioned by only 26 cities.

Weighted Values—It was felt that more weight ought to be given to a listing of a poem "Required for Memory" than to a listing in which the poem was merely "Suggested for Memory," and that still less weight should be given when a poem was suggested for memory while others, presumably better, were required. The least weight would seem to be appropriate when poems were merely "Suggested for Study," although it is to be understood that in this case poems are expected to be much more than merely read by pupils, and parts of them are often recommended for memorizing. Poems merely recommended for reading were omitted from this study entirely. In arriving at a judgment, therefore, of the importance of poems for elementary school use, it was thought proper to give weights of one, two, three, and four respectively to lists made up "For Study," to those made up of poems "Suggested for Memory Others Required," to those made up of poems "Suggested for Memory Without Distinction," and to those made up of poems "Required for Memory." The number of courses of study indicating each type of "mention" was multiplied by its corres-

TABLE II—A SELECTED LIST OF POEMS AND PARTS OF POEMS MENTIONED FIVE OR MORE TIMES IN FIFTY COURSES OF STUDY

No.	Author	Title	Times Mentioned	Cities Mention'g	Weighted Value
1	Smith	America	59	30	207
2	Longfellow	Children's Hour, The	44	39	135
3	Field	Dutch Lullaby	43	41	136
4	Longfellow	Village Blacksmith, The	42	39	134
5	Stevenson	My Shadow	42	40	130
6	Jackson	September	41	36	125
7	Scott	Breathes There a Man	40	38	129
8	Key	Star Spangled Banner	38	26	135
9	Stevenson	Wind, The	38	32	112
10	Browning	Songs from Pippa Passes	35	30	105
11	Hunt	Abou Ben Adhem	35	35	111
12	Jackson	October's Bright Blue Weather	35	32	109
13	Longfellow	Arrow and the Song, The	35	33	116
14	Stevenson	Bed in Summer	35	33	102
15	Lowell	First Snowfall, The	34	26	104
16	Bennett	Flag Goes By, The	33	27	101
17	Hemans	Landing of the Pilgrims, The	33	30	101
18	Ingelow	Seven Times One	33	33	96
19	Tennyson	Sweet and Low	33	31	103
20	Whittier	Barefoot Boy, The	33	31	102
21	Wordsworth	I Wandered Lonely as a Cloud	33	27	96
22	Stevenson	Swing, The	32	31	102
23	Rands	Great, Wide, Beautiful, Wonderful World	32	29	93
24	Field	Night Wind, The	31	29	92
25	Tennyson	Bird and the Baby, The	31	30	94
26	Holmes	Chambered Nautilus, The	30	28	97
27	Child	Thanksgiving Day	29	26	85
28	Field	Why Do Bells for Christmas Ring?	29	25	82
29	Drake	American Flag, The	28	21	94
30	Miller	Blue Bird, The	28	24	81
31	Thaxter	Sandpiper, The	28	21	82
32	Emerson	Concord Hymn	27	24	85
33	Holmes	Old Ironsides	27	27	86
34	Howe	Battle Hymn of the Republic	27	19	94
35	Larcom	Brown Thrush, The	27	25	77
36	Longfellow	Hiawatha's Childhood	27	24	82
37	Sherman	Daisies	27	24	85
38	Stevenson	Land of Story Books, The	27	26	76
39	Field	Little Boy Blue	26	26	76
40	Moore	Visit from St. Nicholas, A	26	24	77
41	Tennyson	Bugle Song	26	24	77
42	Whittier	Corn Song, The	26	23	78
43	Bryant	Robert of Lincoln	25	23	75
44	Longfellow	Day is Done, The	25	23	78
45	Miller	Columbus	25	24	76
46	Tennyson	Brook, The	25	20	73

TABLE II—(*Continued*)

No.	Author	Title	Times Mentioned	Cities Mention'g	Weighted Value
47	Bryant	To a Waterfowl	24	23	76
48	Kipling	Recessional	24	24	76
49	Longfellow	Paul Revere's Ride	24	22	70
50	Shakespeare	Mercy Speech (Merchant of Venice)	24	23	77
51	Bunner	One, Two, Three	23	22	64
52	Carlyle	To-day	23	23	73
53	Cary	November	23	18	63
54	Coolidge	How the Leaves Came Down	23	20	64
55	Hogg	Boy's Song	23	23	67
56	Holland	Gradatim	23	22	71
57	Longfellow	Builders, The	23	21	67
58	Rossetti	Wind, The	23	23	67
59	Taylor	Twinkle, Twinkle, Little Star	23	23	65
60	Whitman	O Captain! My Captain!	23	22	72
61	Brooks	O Little Town of Bethlehem	22	20	67
62	Brown	Little Plant, The	22	22	64
63	Longfellow	Psalm of Life, A	22	20	68
64	Stevenson	Rain	22	22	62
65	Stevenson	Where Go the Boats?	22	21	64
66	Tennyson	Charge of the Light Brigade	22	22	66
67	Emerson	Fable	21	20	75
68	Longfellow	Hiawatha	21	18	56
69	Lowell	Vision of Sir Launfal, The	21	17	61
70	Stevenson	Windy Nights	21	19	57
71	Alexander	All Things Bright and Beautiful	20	20	61
72	Allingham	Wishing	20	19	56
73	Bjornson	Tree, The	20	19	59
74	Cary	Suppose	20	19	60
75	Jackson	Down to Sleep	20	20	58
76	Payne	Home, Sweet Home	20	13	67
77	Stevenson	Cow, The	20	20	54
78	Bryant	Planting of the Apple Tree, The	19	19	75
79	Field	Norse Lullaby	19	19	56
80	Gould	Frost, The	19	18	49
81	Houghton	Good-Night and Good-Morning	19	18	48
82	Houghton	Lady Moon	19	18	56
83	Kingsley	Lost Doll, The	19	19	54
84	Pierpont	Warren's Address	19	17	59
85	Whittier	Snow Bound: A Winter Idyll	19	18	48
86	Aldrich	Marjorie's Almanac	18	17	53
87	Cooper	Come, Litle Leaves	18	16	52
88	Whittier	In School Days	18	17	53
89	Allingham	Fairies, The	17	17	43
90	Cary	Nobility	17	17	53

TABLE II—(*Continued*)

No.	Author	Title	Times Mentioned	Cities Mention'g	Weighted Value
91	Field	Duel, The	17	17	46
92	Finch	Blue and the Gray, The	17	16	51
93	(From the German)	Sleep Baby Sleep	17	16	54
94	MacDonald	Baby, The (At the Back of the North Wind)	17	17	47
95	Riley	Old Glory	17	16	54
96	Rossetti	Boats Sail on the Rivers	17	16	51
97	Thaxter	Spring	17	17	49
98	Allingham	Robin Redbreast	16	16	46
99	Allison	Which Loved Best?	16	16	45
100	Child	Who Stole the Bird's Nest?	16	15	43
101	Field	Japanese Lullaby	16	16	47
102	Field	Rockabye Lady, The	16	16	46
103	Longfellow	Daybreak	16	13	45
104	Longfellow	Hiawatha's Sailing	16	16	48
105	Longfellow	Ship of State, The	16	16	52
106	Morris	Woodman, Spare That Tree	16	15	46
107	Bryant	March	15	14	41
108	Burns	For A' That and A' That	15	15	44
109	Holmes	Last Leaf, The	15	14	45
110	Sherman	Four Winds, The	15	14	42
111	Sill	Opportunity	15	15	42
112	Wadsworth	Over in the Meadow	15	15	39
113	Aldrich	Before the Rain	14	14	43
114	Bryant	Thanatopsis	14	13	39
115	Lear	Owl and the Pussy Cat, The	14	13	43
116	Procter	Sea, The	14	13	38
117	Shakespeare	Under the Greenwood Tree (As You Like It)	14	12	38
118	Shelley	Cloud, The	14	13	36
119	Sherman	Dewdrop, A	14	13	40
120	Stevenson	Autumn Fires	14	14	40
121	Stevenson	Land of Counterpane, The	14	13	47
122	Addison	Spacious Firmament on High, The	13	13	36
123	Bryant	Gladness of Nature, The	13	13	36
124	Bryant	To the Fringed Gentian	13	13	38
125	Cary	Don't Give Up	13	13	40
126	Coleridge	He Prayeth Best	13	12	35
127	Emerson	Snow Storm, The	13	11	37
128	Emerson	We Thank Thee	13	11	33
129	Field	Sugar Plum Tree, The	13	12	37
130	Follen	New Moon, The	13	13	36
131	Longfellow	Rain in Summer	13	12	34
132	Lowell	Day in June, A	13	11	36
133	Lowell	Fatherland, The	13	13	39
134	Stevenson	My Bed is a Boat	13	12	36
135	Stevenson	Sun's Travels, The	13	11	36
136	Tennyson	Break, Break, Break	13	12	40
137	Tennyson	Crossing of the Bar	13	13	40
138	Whittier	Barbara Frietchie	13	13	36

TABLE II—(*Continued*)

No.	Author	Title	Times Mentioned	Cities Mention'g	Weighted Value
139	Coleridge	Answer to a Child's Question	12	10	34
140	Longfellow	Excelsior	12	12	34
141	Lowell	Fountain, The	12	11	30
142	MacDonald	Wind and the Moon, The	12	12	35
143	Riley	Sudden Shower, A	12	10	28
144	Shakespeare	Polonius to Laertes (Hamlet)	12	11	36
145	Stevenson	Lamplighter, The	12	12	31
146	Taylor	I Like Little Pussy	12	12	33
147	Taylor	Violet, The	12	12	34
148	Wordsworth	To a Butterfly	12	11	42
149	Brooks	Christmas Everywhere	11	10	32
150	Browning	Incident of the French Camp, An	11	11	29
151	Fields	Ballad of the Tempest, The	11	11	30
152	Hopkinson	Hail Columbia!	11	4	31
153	Longfellow	Building of the Ship, The	11	10	29
154	Read	Sheridan's Ride	11	11	35
155	Riley	Little Orphant Annie	11	11	30
156	Sherman	Golden Rod	11	9	32
157	Tennyson	Owl, The	11	10	29
158	Thaxter	March	11	11	24
159	Blake	Lamb, The	10	10	31
160	Holmes	Union and Liberty	10	10	37
161	Hugo	Good Night	10	10	30
162	Kingsley	Farewell, A	10	10	31
163	Krout	Little Brown Hands	10	10	28
164	Longfellow	Rainy Day, The	10	10	28
165	Lowell	To the Dandelion	10	8	25
166	Newman	Lead Kindly Light	10	10	31
167	Perry	Coming of Spring, The	10	10	27
168	Riley	Song, A	10	10	27
169	Shelley	Sky Lark, The	10	10	28
170	Sherman	Leaves at Play	10	10	27
171	Stevenson	Foreign Children	10	9	25
172	Stevenson	Foreign Lands	10	10	28
173	Tate	Christmas	10	9	23
174	Tennyson	Ring Out, Wild Bells	10	10	30
175	Wordsworth	Written in March	10	10	26
176	Browning	Child's Thought of God, A	9	9	23
177	Bryant	Death of the Flowers, The	9	9	25
178	Bryant	Song of Marion's Men	9	9	23
179	Cone	Dandelions, The	9	8	26
180	Garabrant	Dandelion	9	9	24
181	Gray	Elegy Written in a Country Churchyard	9	9	22
182	Jewett	Discontent	9	9	24
183	Kipling	White Seal, The	9	9	24
184	Macauley	Horatius	9	9	18
185	Riley	Knee Deep in June	9	9	27
186	Shakespeare	Hark, Hark! the Lark (Cymbeline)	9	9	23

TABLE II—(*Continued*)

No.	Author	Title	Times Mentioned	Cities Mention'g	Weighted Value
187	Sherman	Clouds	9	9	25
188	Stedman	What the Winds Bring	9	9	24
189	Thaxter	Wild Geese	9	9	20
190	Thomas	Talking in Their Sleep	9	9	24
191	Vandegrift	Sand Man, The	9	9	24
192	Wolfe	Burial of Sir John Moore	9	9	27
193	Wordsworth	To a Skylark	9	9	23
194	Browning	How They Brought the Good News	8	8	18
195	Campbell	Hohenlinden	8	8	22
196	Edwards	Child's Prayer, A	8	8	21
197	Field	Seein' Things	8	7	18
198	Herford	Elf and the Dormouse, The	8	7	17
199	Higinson	Four Leaf Clovers	8	8	20
200	Howitt	Voice of Spring, The	8	8	17
201	Larcom	If I Were a Sunbeam	8	8	24
202	Longfellow	Bell of Atri, The	8	8	18
203	Longfellow	Evangeline	8	8	19
204	Longfellow	Old Clock on the Stairs, The	8	8	22
205	Longfellow	Wreck of the Hesperus, The	8	8	20
206	Lowell	Heritage, The	8	8	21
207	Lowell	Yussouf	8	8	23
208	Riley	Brook Song, The	8	8	22
209	Scott	Lochinvar	8	7	21
210	Shakespeare	Good Name (Othello)	8	8	23
211	Sherman	Snowflakes	8	8	17
212	Sherman	Wizard Frost	8	7	17
213	Southey	Inchcape Rock, The	8	7	26
214	Taylor	Thank You Pretty Cow	8	8	16
215	Tennyson	Flower in the Crannied Wall	8	8	22
216	Tennyson	Throstle, The	8	8	21
217	Thaxter	Little Gustava	8	8	20
218	Wordsworth	My Heart Leaps up When I Behold	8	8	19
219	Bates	Who Likes the Rain?	7	7	16
220	Browning	Home Thoughts from Abroad	7	6	17
221	Cary	Order for a Picture, An	7	7	19
222	Cary	They Didn't Think	7	7	19
223	Cooper	October's Party	7	7	22
224	Cooper	Wonderful Weaver, The	7	6	15
225	Cooper	What Robin Told	7	7	21
226	Holland	Christmas Carol, A	7	7	20
227	Hood	I Remember, I Remember	7	7	17
228	Kipling	L'Envoi	7	7	19
229	Lanier	Song of the Chattahoochee, The	7	7	18
230	Lowell	Aladdin	7	7	20
231	Lowell	Finding of the Lyre, The	7	7	17
232	Mackay	Tubal Cain	7	7	17
233	Riley	Life Lesson, A	7	7	22

TABLE II—(*Continued*)

No.	Author	Title	Times Mentioned	Cities Men-tion'g	Weighted Value
234	Scott	Lullaby of an Infant Chief	7	7	17
235	Shakespeare	Mark Anthony's Address (Julius Caesar)	7	7	21
236	Shakespeare	Orpheus with His Lute (King Henry the Eighth)	7	7	17
237	Sherman	Real Santa Claus, A	7	7	15
238	Sherman	Hide and Seek	7	7	14
239	Sherman	Snowbird, The	7	7	15
240	Taylor	Song of the Camp, The	7	7	18
241	Tennyson	In Memoriam	7	6	21
242	Tennyson	King Arthur	7	7	20
243	Tennyson	Sir Galahad	7	7	17
244	Tennyson	Window, The	7	7	18
245	Watts	Busy Bee, The	7	7	15
246	Whittier	Huskers, The	7	5	15
247	Wordsworth	Lucy Gray	7	7	17
248	Wordsworth	Kitten and the Falling Leaves, The	7	7	17
249	Anonymous	Laughing Chorus, A	6	6	16
250	Bangs	Little Elf, The	6	6	13
251	Blake	Tiger, The	6	6	17
252	Burns	Bannockburn	6	6	14
253	Byron	Eve of Waterloo, The	6	6	16
254	Campbell	Lord Ullin's Daughter	6	6	14
255	Cary	Leak in the Dyke, The	6	6	15
256	Emerson	Rhodora, The	6	6	13
257	Hale	Mary's Lamb	6	6	15
258	Hemans	Casabianca	6	6	12
259	Hogg	Skylark, The	6	6	15
260	Howlister	Our Flag	6	6	21
261	Jones	What Constitutes a State?	6	6	15
262	Kingsley	Three Fishers, The	6	6	14
263	Kipling	If	6	5	19
264	Lang	Scythe Song	6	6	14
265	Lear	Nonsense Alphabet	6	5	13
266	Longfellow	Courtship of Miles Standish, The	6	6	12
267	Longfellow	Ladder of St. Augustine, The	6	6	17
268	Lowell	Stanzas on Freedom	6	6	17
269	McDonald	Little White Lily, The	6	6	16
270	Mackay	Miller of the Dee, The	6	6	18
271	Nesbitt	Your Flag and My Flag	6	6	20
272	Proctor	One by One	6	6	16
273	Rossetti	Milking Time	6	6	15
274	Sherman	May	6	5	15
275	Stevenson	Happy Thought	6	6	15
276	Stevenson	Whole Duty of Children	6	6	18
277	Thaxter	Chanticleer	6	6	12
278	Wilder	Stand by the Flag	6	6	16
279	Wordsworth	We Are Seven	6	5	12
280	Alexander	Burial of Moses, The	5	5	14
281	Anonymous	Secret, The	5	5	13
282	Browning	Pied Piper of Hamelin, The	5	5	9

TABLE II—(*Continued*)

No.	Author	Title	Times Mentioned	Cities Mention'g	Weighted Value
283	Bryant	Yellow Violet, The	5	5	11
284	Burns	My Heart's in the Highlands	5	5	13
285	Byron	Destruction of Sennacherib	5	5	13
286	Cary	Obedience	5	5	13
287	Cowper	Nightingale and the Glowworm, The	5	5	12
288	Deland	While Shepherds Watched Their Flocks by Night	5	5	15
289	Dodge	Snowflakes	5	5	12
290	Emerson	Each and All	5	5	12
291	Emerson	Forbearance	5	5	11
292	Howitt	Fairies of the Caldon Low, The	5	5	10
293	Jelliffe	Clovers, The	5	5	11
294	Keats	On the Grasshopper and Cricket	5	5	12
295	Keble	All Things Beautiful	5	5	13
296	Larcom	Calling the Violet	5	5	12
297	Larcom	Rivulet, The	5	5	12
298	Longfellow	Hiawatha's Friends	5	5	13
299	Longfellow	King Robert of Sicily	5	5	11
300	Longfellow	Sandalphon	5	5	11
301	Lowell	Ode Recited at the Harvard Commemoration	5	5	13
302	Milton	On His Blindness	5	5	15
303	Milton	Song on a May Morning	5	5	13
304	Moore	Minstrel Boy, The	5	5	13
305	Poulsson	First Christmas, The	5	5	14
306	Poulsson	While Stars of Christmas Shine	5	5	18
307	Rossetti	O Lady Moon	5	5	14
308	Rossetti	Swallow, The	5	5	13
309	Rossetti	What Does the Bee Do?	5	5	12
310	Sangster	We Thank Thee	5	5	18
311	Saxe	Blind Men and the Elephant, The	5	5	14
312	Scott	Lady of the Lake	5	5	7
313	Shakespeare	Ariel's Song (The Tempest)	5	5	12
314	Shaw	Columbia, the Gem of the Ocean	5	5	18
315	Sherman	Song for Winter	5	5	8
316	Southey	Little Ladybird, The	5	4	8
317	Stevenson	Good Play, A	5	5	12
318	Stevenson	Farewell to the Farm	5	5	16
319	Stevenson	Singing	5	5	11
320	Stevenson	Moon, The	5	5	15
321	Stevenson	Time to Rise	5	5	13
322	Stoddard	Abraham Lincoln	5	5	14
323	Tennyson	Eagle, The	5	5	13
324	Thaxter	April	5	5	12
325	Thaxter	Piccola	5	5	11
326	Van Dyke	Song Sparrow, The	5	5	12
327	Whittier	Three Bells, The	5	5	9
328	Wordsworth	Fidelity	5	4	9
329	Wordsworth	Pet Lamb, The	5	5	11

ponding weight and the sum of the products was taken to give the "Weighted Value" appearing in the last column of Table II. For instance, "America" was mentioned as follows:

	Frequency	Weight	Product
For Study	1	1	1
Suggested for Memory Others Required	3	2	6
Suggested for Memory without Distinction	20	3	60
Required for Memory	35	4	140
TOTAL	59		207

Weighting these frequencies of mention according to our scheme we have for the weighted value of "America" 207 as shown in Table II. This value greatly exceeded that of any other poem. Poems which were mentioned an equal number of times might thus vary considerably in weighted value according to the type of mention which they received. For example, "Abou Ben Adhem," "The Arrow and the Song," and "Bed in Summer," were all mentioned 35 times, but varied in weighted value from 102 to 116. "Bed in Summer" and "Abou Ben Adhem" were definitely required for memory only 5 and 9 times respectively; while "The Arrow and the Song," which received the weighted value of 116 was required for memory by 12 cities.

Correspondence in Ranking by the Three Methods—An examination of Table II will show in a general way the extent to which the ranking of the poems is the same under the three captions "Times Mentioned," "Cities Mentioning," and "Weighted Value." The table is arranged according to the decreasing frequencies of "Times Mentioned." Considering only the first twenty-five poems (those mentioned 31 or more times) one finds that but two of them would fail to be included in a list of the twenty-five poems mentioned by the largest number of cities. In other words, the twenty-five most popular poems according to the one arrangement are also the twenty-five most popular poems according to the other arrangement with but two exceptions. These poems are "The Star Spangled Banner" and Lowell's "The First Snowfall." Although they have a ranking of 8 and 15 respectively on the basis of "Times Mentioned" they were listed in each case by but 26 *different* cities.

Again, only 3 poems (Holmes' "Chambered Nautilus"; Drake's "The American Flag"; and Howe's "The Battle Hymn of the Republic") which fall outside the first 25 poems according to "Times Mentioned" are entitled to a ranking among the first 25 on the basis of "Weighted Value." Here as in so many other instances, the introduction of weights proves to have little practical utility—at least where the number of measures is fairly large.

A general expression of the closeness of correspondence between the three arrangements is furnished for the entire 329 poems by the correlation coefficients. These are:

1. Correlation between "Times Mentioned" and "Cities Mentioning"	0.96
2. Correlation between "Times Mentioned" and "Weighted Value"	0.99
3. Correlation between "Cities Mentioning" and "Weighted Value"	0.95

Evidently either one of these bases of arrangements will yield much the same results. Only occasionally, as in the cases pointed out above, does this fail to be the case, and even then the deviations are not large.

At this point it would be natural to mention the ten, twenty, thirty, etc., most frequently used poems. Since, however, a reading of Table II at once discloses this information, it is clearly unnecessary to repeat here the names and authors of such poems. Those interested in selecting for school use, the poems which are now used to the greatest extent throughout the country may do so directly from Table II.

Comparison of the Selected List with Atherton's List—As a criticism of the Selected List we have compared it with the list submitted by Atherton in the article mentioned above. His list consisted of 382 titles; ours of 329. There were 255 titles common to both lists. Therefore, 66.7% or two-thirds of the 382 poems in Atherton's list appeared in the "Selected List." Similarly, 77.5% or more than three-fourths of the poems in the Selected List, were also included in his list. If the "Selected List" had contained all the poems mentioned 3 or more times (as Atherton's list did) practically all of the selections in his list would have been included. Since the courses of study used in the two investigations covered different periods of time and for the most part different localities, it is clear that the "Selected List" contains poems which are considered by concensus of opinion to be the most valuable for memory work in the schools. Moreover, since the courses of study used by Atherton's investigation preceded those used in ours by from five to ten years, it would appear that material chosen for memory work has scarcely changed during recent years. Poems which have found a place in courses of study are those which have stood the test of time. Teachers will find it necessary to supplement such lists by more recent poetry which in their judgment should be included.

The Newer Poetry—A few of the 1919 courses of study did include some of the best-known war poems. McRae's "In Flanders Fields," Service's "Carry On," and Seegar's "Rendezvous with Death" were all noted. One course of study required for memory "Keep the Home Fires Burning" and "There's A Long, Long Trail A-Winding." The inclusion of the last two selections illustrates the recent tendency of some courses to include almost any material dealing with the war regardless of its value. No recent poems other than those dealing with the war were listed by any of the courses of study.

If we may comment in general terms on this condition, we should say that in this type of material we have a conspicuous example of the extent to which the course of study in our schools lags behind current conditions. We pay a price for the assurance that each piece of literary material has "stood the test of time."

V—Grading the Selected List

One of the most important steps in the study of standards in memory work is to determine the grades in which the various poems should be taught. Table III shows a grading of the 329 poems in the Selected List. The figures in the first column correspond to those preceding the poems in Table II. Thus poem No. 1 in Table III is "America"; poem No. 2 is "The Children's Hour"; etc. The figures in the body of the table indicate the number of times each poem was mentioned in grades one to eight inclusive. The figures in the column headed "Total" give the total number of times each poem was mentioned. They are the same as the entries in the column headed "Times Mentioned" in Table I.

How the Poems Were Graded—The grade in which a poem received the highest frequency of mention was, in general, taken as the grade to which, on the basis of practice, it belonged, the argument being that the most reasonable grading was that agreed to by the greatest number. Poem 1 was mentioned fifteen times in Grade II. This is a higher frequency than was found for it in any other grade; therefore Grade II is the one to which usage assigns "America" for memorizing. Similarly, No. 2 with a frequency of 20 in Grade III is assigned to that grade. Occasionally a poem was found with an equal frequency in two adjacent grades. No. 54 (Coolidge's "How the Leaves Came Down") is an example of this case. The poem was mentioned eight times in Grade II and the same number of times in Grade III. However, it was mentioned six times in Grade I and only once in Grade IV. Therefore, the greater frequency in Grade I suggests that the poem is better placed in Grade II than in Grade III.

No. 259 (Hogg's "The Skylark") shows a frequency of two both in Grade IV and in Grade VI. It was not mentioned in Grade V by any course of study, but the poem was mentioned once in Grades III and VIII. It was, however, placed in the fifth grade because just as many schools had seen fit to place it above the fifth as had placed it below the fifth. If a greater number of schools had mentioned the poem, there probably would have been a piling up of frequency in Grade V.

Patriotic poems show as a rule a fairly even distribution in all the grades. This condition is largely due to the fact that many cities require these poems throughout the school system. Nevertheless, the grade receiving the highest frequency is taken as the grade in which the poem may be most appropriately taught. Patriotic reasons may dictate the presentation of such poems in earlier grades, but the children in these grades are unlikely to appreciate their thought. It must be remembered that although our patriotic poetry appeals to a relatively simple emotion, it was not composed for children and the style in which it is written is a mature style. One needs but to run over in mind the first stanzas of "America" or of the "Star Spangled Banner" with their inversions and involved constructions, to appreciate the fact that children of the lower grades rarely understand their meaning.

TABLE III—A SUGGESTED GRADING OF POEMS MENTIONED FIVE OR MORE TIMES

No.	Number of Times Mentioned in the Indicated Grades								Total	Grade Suggested
	I	II	III	IV	V	VI	VII	VIII		
1	11	15	10	11	4	4	2	2	59	II
2	3	5	20	12	4				44	III
3	6	22	13	2					43	II
4		3	6	20	12	1			42	IV
5	21	20	1						42	I
6	4	11	16	10					41	III
7			2	2	4	10	7	15	40	VIII
8	2	2	3	6	11	5	7	2	38	V
9	14	16	6		1			1	38	II
10	4	5	7	4	6	4	5		35	V
11				1	12	15	4	3	35	VI
12		2	10	14	7	2			35	IV.
13	2	1	2	11	10	2	6	1	35	IV
14	22	12	1						35	I
15		1	4	11	9	8	1		34	IV
16		2	3	6	10	12			33	V
17			2	4	13	11	2	1	33	V
18	4	18	11						33	II
19	3	6	14	7	2			1	33	III
20		2	2	10	15	4			33	V
21			1	10	5	10	1	6	33	VI
22	23	7	2						32	I
23	1	9	17	5					32	III
24		11	1	17	2				31	IV
25	29	2							31	I
26					1	2	10	17	30	VIII
27	5	12	9	2		1			29	II
28	14	12	3						29	I
29	1	1	4	1	2	8	8	3	28	VI
30	2	13	7	6					28	II
31			4	19	4	1			28	IV
32				1	4	8	6	8	27	VI
33				1	3	14	4	5	27	VI
34	1	1	2	2	3	3	5	10	27	VIII
35		15	10		2				27	II
36	6	16	5						27	II
37	10	15	2						27	II
38	3	12	11	1					27	II
39	9	7	6	3				1	26	I
40	6	3	14	2	1				26	III
41			1	1	2	11	6	5	26	VI
42			1	8	11	6			26	V
43	1		5	11	4	3		1	25	IV
44				1	14	8		2	25	V
45				1	2	8	11	3	25	VII
46			7	7	5	3	1	2	25	IV
47					2	6	8	8	24	VII
48							5	19	24	VIII
49			1	3	9	7	3	1	24	V
50					1	2	4	17	24	VIII
51	4	9	8	2					23	II

TABLE III (*Continued*)

No.	Number of Times Mentioned in the Indicated Grades								Total	Grade Suggested
	I	II	III	IV	V	VI	VII	VIII		
52			3	1	13	4	2		23	V
53	3	6	9	1	3		1		23	III
54	6	8	8	1					23	II
55	1	1	11	10					23	III
56				2	4	5	7	5	23	VII
57					6	14	3		23	VI
58	20	3							23	I
59	21	2							23	I
60							6	17	23	VIII
61	1		8	5	5	3			22	III
62	21		1						22	I
63					6	13	2	1	22	VI
64	20	2							22	I
65	6	9	7						22	II
66					2	10	9	1	22	VI
67			3	3	11	4			21	V
68	3	6	4	3	2	1	1	1	21	II
69			1	1	1	2	5	11	21	VIII
70	7	10	3		1				21	II
71	10	8	1	1					20	I
72		5	12	3					20	III
73	2	6	11	1					20	III
74	2	9	8	1					20	II
75			3	3	10	4			20	V
76	1	1	1	6	5	3	1	2	20	IV
77	19	1							20	I
78			1	5	10	1	2		19	V
79	1	2	11	5					19	III
80	1	5	5	7	1				19	IV
81	3	8	8						19	II
82	9	9	1						19	II
83	8	8	2	1					19	II
84					3	3	6	7	19	VIII
85			1	1	2	2	5	8	19	VIII
86		4	12	2					18	III
87	10	7	1						18	I
88		1	1	3	8	5			18	V
89		7	5	4		1			17	II
90				2	2	10	3		17	VI
91	2	13	2						17	II
92				1	1	9	3	3	17	VI
93	15	2							17	I
94	14	2	1						17	I
95			2	1	3	5	3	3	17	VI
96	14	3							17	I
97	3	8	6						17	II
98	2	8	5	1					16	II
99	6	6	4						16	II
100	7	7	2						16	II
101	1	6	4	5					16	II
102	5	7	4						16	II
103		2	2	6	3	1	2		16	IV

TABLE III—(*Continued*)

No.	Number of Times Mentioned in the Indicated Grades								Total	Grade Suggested
	I	II	III	IV	V	VI	VII	VIII		
104		2	11	1	2				16	III
105				1	6	3	2	4	16	V
106			2	3	9	2			16	V
107			3	2	3	7			15	VI
108					1	4	2	8	15	VIII
109				1	2	2	3	7	15	VIII
110	2	7	4		2				15	II
111						1	4	10	15	VIII
112	14	1							15	I
113					3	11			14	VI
114							4	10	14	VIII
115	2	5	5	2					14	II
116			1	5	3	3		2	14	IV
117				4	6	1		3	14	V
118		1	1	2		2	3	5	14	VIII
119	10	3	1						14	I
120	8	5	1						14	I
121	6	4	4						14	I
122				1		4	5	3	13	VII
123				2	8		3		13	V
124			1	1	3	3	3	2	13	VI
125	1		12						13	III
126			5	6		1	1		13	IV
127			2	1	1	3	5	1	13	VII
128	4	7	1	1					13	II
129	6	5	1		1				13	I
130	6	7							13	II
131			2	6	4	1			13	IV
132				2	1	1	2	7	13	VIII
133			1	1	2	2	4	3	13	VII
134	5	6	2						13	II
135	3	8	2						13	II
136			1		3	7	1	1	13	VI
137					1	2	5	5	13	VII
138				2	9	1	1		13	V
139		6	5	1					12	II
140				2	8	1	1		12	V
141			3	6	1	1		1	12	IV
142	1	1	2	6	2				12	IV
143			4	5	2	1			12	IV
144							1	11	12	VIII
145	6	3	3						12	I
146	11	1							12	I
147		2	8	2					12	III
148		1	7	3		1			12	III
149	1	1	1	1	3	4			11	VI
150					2	3	4	2	11	VII
151		1	5	2	2	1			11	III
152	1	1	2	2	1	2	1	1	11	IV
153						1	7	3	11	VII
154					1	6	2	2	11	VI
155			7	2	1	1			11	III

TABLE III—(*Continued*)

No.	NUMBER OF TIMES MENTIONED IN THE INDICATED GRADES								Total	Grade Suggested
	I	II	III	IV	V	VI	VII	VIII		
156	4	3	1	3					11	I
157		1	7	3					11	III
158	2	1	4	2	1	1			11	III
159	5	3	1	1					10	I
160						3	5	2	10	VII
161	9		1						10	I
162			2	5	2	1			10	IV
163			5	3	2				10	III
164			1	2	1	1	4	1	10	VII
165			2	1		5		2	10	VI
166					3	1	2	4	10	VIII
167		2	1	1	2	4			10	VI
168			3	1	3	1	2		10	V
169					1		3	6	10	VIII
170	5	4	1						10	I
171	2	5	2	1					10	II
172	1	4	4	1					10	II
173			3	3	1	3			10	IV
174					2	4	2	2	10	VI
175		2	2	4	2				10	IV
176		1	6	2					9	III
177				1		3	2	3	9	VII
178						2	5	2	9	VII
179			5	1	1	1	1		9	III
180	1	4		3		1			9	II
181							1	8	9	VIII
182	1	5	1	2					9	II
183		3		2	2	1	1		9	II
184			1			2	4	2	9	VII
185			1	1	3	3	1		9	V
186			1	3		1	3	1	9	VI
187	1	1	5	2					9	III
188	1	2	6						9	III
189	1	4	2	2					9	II
190	1	1	4	3					9	III
191		1	8						9	III
192				1		5	1	2	9	VI
193					1	5	3		9	VI
194						4	3	1	8	VI
195					1	3	4		8	VII
196	2	2	4						8	III
197		1	4	2	1				8	III
198	1	6	1						8	II
199	1	1	1	2	3				8	V
200			5	2	1				8	III
201		7	1						8	II
202				4	1	2	1		8	IV
203							5	3	8	VII
204	1			1	5	1			8	V
205			2	4	2				8	IV
206				1	3	2	2		8	V

TABLE III—(*Continued*)

No.	Number of Times Mentioned in the Indicated Grades								Total	Grade Suggested
	I	II	III	IV	V	VI	VII	VIII		
207				2	1	1	2	2	8	VII
208				7	1				8	IV
209					1	3	2	2	8	VI
210				1	1		4	2	8	VII
211	3	2	3						8	II
212		3	4	1					8	III
213				1	4	1	1	1	8	V
214	5	2	1						8	I
215					2	1	2	3	8	VIII
216	2		3	1		1		1	8	III
217	2	2	3		1				8	III
218		2	1	1	2	1		1	8	V
219	4	2			1				7	I
220						1	4	2	7	VII
221					1	2	1	3	7	VI
222	5	1	1						7	I
223	3	3		1					7	II
224		1	5	1					7	III
225	1	5		1					7	II
226	1		2			3		1	7	VI
227			2	3		1	1		7	IV
228							2	5	7	VIII
229					1	1	1	4	7	VIII
230					4	2	1		7	V
231						2	5		7	VII
232			1		2	2	1	1	7	VI
233					2	1	3	1	7	VII
234		3	1	1	2				7	II
235								7	7	VIII
236					2	5			7	VI
237		3	4						7	III
238	3	4							7	II
239	4	3							7	I
240					1	2	3	1	7	VII
241					1	1	3	2	7	VII
242						2	2	3	7	VIII
243							4	3	7	VII
244	2		1	2	2				7	IV
245	4	3							7	I
246			1		1	2	3		7	VII
247			3	3			1		7	IV
248			5	2					7	III
249	2		1	2	1				6	IV
250	1	5							6	II
251				2	1	2		1	6	VI
252					1		3	2	6	VII
253							1	5	6	VIII
254						5		1	6	VI
255			1	1	2	2			6	V
256						2	2	2	6	VII
257	6								6	I
258			2	2	2				6	IV

TABLE III—(*Continued*)

No.	NUMBER OF TIMES MENTIONED IN THE INDICATED GRADES								Total	Grade Suggested
	I	II	III	IV	V	VI	VII	VIII		
259			1	2		2		1	6	V
260		3	2	1					6	II
261							2	4	6	VIII
262				1	1	2	1	1	6	VI
263						1	2	3	6	VIII
264						4	1	1	6	VI
265	4	1			1				6	I
266						1	5		6	VII
267						2	2	2	6	VII
268							1	5	6	VIII
269	2	4							6	II
270			3		1	2			6	III
271	1		3		1	1			6	III
272			1		2	3			6	VI
273		4	2						6	II
274		1		1	1	3			6	VI
275	6								6	I
276	5	1							6	I
277	2	2	1	1					6	II
278					2	1	3		6	VII
279			2	2	1	1			6	IV
280					1	1	2	1	5	VII
281	5								5	I
282				1		2	1	1	5	VI
283			1	1	1	2			5	VI
284					2		3		5	VII
285						1	1	3	5	VIII
286		4	1						5	II
287			3		1	1			5	III
288			2	2	1				5	IV
289	2	3							5	II
290					1		1	3	5	VIII
291				2	2		1		5	V
292		1	3	1					5	III
293	1	4							5	II
294					3	1		1	5	V
295	3	2							5	I
296		1	1	3					5	IV
297		1	4						5	III
298		1	1	2	1				5	IV
299						1	2	2	5	VII
300						3		2	5	VI
301								5	5	VIII
302						1		4	5	VIII
303					3		1	1	5	V
304				1		3	1		5	VI
305	3	1	1						5	I
306	3	1	1						5	I
307	4	1							5	I
308	1	1	2	1					5	III
309	1	1							5	I
310	1	3		1					5	II

TABLE III—(*Continued*)

No.	Number of Times Mentioned in the Indicated Grades								Total	Grade Suggested
	I	II	III	IV	V	VI	VII	VIII		
311				1	2	2			5	V
312							1	4	5	VIII
313	1	2		2					5	II
314		1	3	1					5	III
315	3	2							5	I
316	2		2	1					5	III
317	2	2	1						5	II
318		4	1						5	II
319	5								5	I
320	2	1	1	1					5	I
321	4	1							5	I
322					1		2	2	5	VII
323		1		1		1	1	1	5	VI
324	2	2		1					5	II
325	1	1			2	1			5	V
326	1		2		1	1			5	III
327				3	2				5	IV
328					1	3	1		5	VI
329	1	1		2	1				5	IV

The Selected List Arranged by Grades—Tables IV to XI group the 329 poems in the Selected List according to the grades to which the agreements of the courses of study suggest that they belong. The poems are arranged alphabetically by authors. The column headed "Frequency" indicates the number of times each poem was named in the grade to which it is assigned. The tables also show the "percent of frequency" with which a poem was mentioned in the given grade based on the total frequency of mention which it received for all grades. Other things being equal poems which receive a high "percent of frequency" in a given grade may be chosen as poems best suited for memory work in that grade. In selecting poems for memory lists from Tables IV to XI the teacher should also give attention to the total frequency of mention (the index of importance of the poem) as well as to the "percent of total frequency" for the grade in question. For example, Alexander's "All Things Bright and Beautiful" was mentioned ten times for memorizing by children of Grade I. It received a total frequency of mention of 20 and was, therefore, assigned by 50 percent of the courses as best suited to Grade I. Similarly, Brown's "The Little Plant" was mentioned 21 times and shows a "percent of total frequency" of 95.4. The total frequency may be found by referring to Tables II and III. In order to facilitate finding the poems in Tables II and III the code numbers of the poems are indicated in the last column of Tables IV to XI.

TABLE IV—LIST OF POEMS FOR MEMORIZING IN GRADE I

Author	Title	Frequency in Grade I	Per Cent of Total Frequency	Code No.
Alexander	All Things Bright and Beautiful	10	50.0	71
Anonymous	Secret, The	5	100.0	281
Anonymous*	Sleep, Baby, Sleep	15	88.2	93
Bates	Who Likes the Rain?	4	59.1	219
Blake	Lamb, The	5	50.0	159
Brown	Little Plant, The	21	95.4	62
Cary, P.	They Didn't Think	5	71.3	222
Cooper	Come, Little Leaves	10	55.5	87
Field	Little Boy Blue	9	34.6	39
Field	Sugar Plum Tree, The	6	46.2	129
Field	Why Do Bells for Christmas Ring?	14	48.3	28
Hale	Mary's Lamb	6	100.0	257
Hugo	Good Night	9	90.0	161
Keble	All Things Beautiful	3	60.0	295
Lear	Nonsense Alphabet	4	66.7	265
MacDonald	At the Back of the North Wind	14	82.4	94
Poulsson	First Christmas, The	3	60.0	305
Poulsson	While Stars of Christmas Shine	3	60.0	306
Rossetti	Boats Sail on the Rivers	14	82.4	96
Rossetti	O, Lady Moon	4	80.0	307
Rossetti	What Does the Bee Do?	4	80.0	309
Rossetti	Wind, The	20	86.9	58
Sherman	Dewdrop, A	10	71.4	119
Sherman	Golden Rod	4	36.4	156
Sherman	Leaves at Play	5	50.0	170
Sherman	Snowbird, The	4	57.1	239
Sherman	Song for Winter	3	60.0	315
Stevenson	Autumn Fires	8	57.1	120
Stevenson	Bed in Summer	22	62.9	14
Stevenson	Cow, The	19	95.0	77
Stevenson	Happy Thought	6	100.0	275
Stevenson	Lamplighter, The	6	50.0	145
Stevenson	Land of Counterpane, The	6	42.9	121
Stevenson	Moon, The	2	40.0	320
Stevenson	My Shadow	21	50.0	5
Stevenson	Rain	20	90.9	64
Stevenson	Singing	5	100.0	319
Stevenson	Swing, The	23	71.9	22
Stevenson	Time to Rise	4	80.0	321
Stevenson	Whole Duty of Children	5	83.3	276
Taylor	I Like Little Pussy	11	91.7	146
Taylor	Thank You, Pretty Cow	5	62.5	214
Taylor	Twinkle, Twinkle, Little Star	21	91.3	59
Tennyson	Bird and the Baby, The	29	93.6	25
Wadsworth	Over in the Meadow	14	93.3	112
Watts	Busy Bee, The	4	55.6	245

*From the German

TABLE V—LIST OF POEMS FOR MEMORIZING IN GRADE II

Author	Title	Frequency in Grade II	Per Cent of Total Frequency	Code No.
Allingham	Fairies, The	7	41.2	89
Allingham	Robin Redbreast	8	50.0	98
Allison	Which Loved Best?	6	37.5	99
Bangs	Little Elf, The	5	83.3	250
Bunner	One, Two, Three	9	39.1	51
Cary, P.	Obedience	4	80.0	286
Cary, P.	Suppose	9	45.0	74
Child	Thanksgiving Day	12	41.4	27
Child	Who Stole the Bird's Nest?	7	43.8	100
Coleridge	Answer to a Child's Question	6	50.0	139
Coolidge	How the Leaves Came Down	8	34.9	54
Cooper	October's Party	3	42.8	223
Cooper	What Robin Told	5	71.4	225
Dodge	Snowflakes	3	60.0	289
Emerson	We Thank Thee	7	53.8	128
Field	Duel, The	13	76.5	91
Field	Dutch Lullaby	22	51.2	3
Field	Japanese Lullaby	6	37.5	101
Field	Rockabye Lady, The	7	43.8	102
Follen	New Moon, The	7	53.8	130
Garabrant	Dandelion	4	44.4	180
Herford	Elf and the Dormouse, The	6	75.0	198
Houghton	Good-Night and Good-Morning	8	42.1	81
Houghton	Lady Moon	9	47.4	82
Howlister	Our Flag	3	50.0	260
Ingelow	Seven Times One	18	54.5	18
Jelliffe	Clovers, The	4	80.0	293
Jewett	Discontent	5	55.5	182
Kingsley	Lost Doll, The	8	42.1	83
Kipling	White Seal, The	3	33.3	183
Larcom	If I Were a Sunbeam	7	87.5	201
Larcom	Brown Thrush, The	15	55.5	35
Lear	Owl and the Pussy Cat, The	5	35.7	115
Longfellow	Hiawatha	6	28.6	68
Longfellow	Hiawatha's Childhood	16	59.3	36
MacDonald	Little White Lily, The	4	66.7	269
Miller	Blue Bird, The	13	46.4	30
Rossetti	Milking Time	4	66.7	273
Sangster	We Thank Thee	3	60.0	310
Scott	Lullaby of an Infant Chief	3	42.9	234
Shakespeare	Ariel's Song (Tempest, The)	2	40.0	313
Sherman	Daisies	15	55.5	37
Sherman	Four Winds, The	7	46.7	110
Sherman	Hide and Seek	4	57.1	238
Sherman	Snowflakes	2	25.0	211
Smith	America	15	25.4	1
Stevenson	Good Play, A	2	40.0	317
Stevenson	Farewell to the Farm	4	80.0	318
Stevenson	Foreign Children	5	50.0	171
Stevenson	Foreign Lands	4	40.0	172
Stevenson	Land of Story Books, The	12	44.4	38
Stevenson	My Bed is a Boat	6	46.2	134
Stevenson	Sun's Travels, The	8	61.5	135
Stevenson	Where Go the Boats	9	40.9	65

TABLE V—(*Continued*)

Author	Title	Frequency in Grade II	Per Cent of Total Frequency	Code No.
Stevenson	Wind, The	16	42.1	9
Stevenson	Windy Nights	10	47.6	70
Thaxter	April	2	40.0	324
Thaxter	Chanticleer	2	33.3	277
Thaxter	Spring	8	46.5	97
Thaxter	Wild Geese	4	44.4	189

TABLE VI—LIST OF POEMS FOR MEMORIZING IN GRADE III

Author	Title	Frequency in Grade III	Per Cent of Total Frequency	Code No.
Aldrich	Marjorie's Almanac	12	66.7	86
Allingham	Wishing	12	60.0	72
Bjornson	Tree, The	11	55.5	73
Brooks	O Little Town of Bethlehem	8	36.4	61
Browning	Child's Thought of God, A	6	66.7	176
Cary, A.	November	9	39.1	53
Cary, P.	Don't Give Up	12	92.3	125
Cone	Dandelions, The	5	55.5	179
Cooper	Wonderful Weaver, The	5	71.4	224
Cowper	Nightingale and the Glow-worm, The	3	60.0	287
Edwards	Child's Prayer, A	4	50.0	196
Field	Norse Lullaby	11	57.9	79
Field	Seein' Things	4	50.0	197
Fields	Ballad of the Tempest, The	5	45.5	151
Hogg	Boy's Song	11	47.8	55
Howitt	Fairies of the Caldon Low, The	3	60.0	292
Howitt	Voice of Spring, The	5	62.5	200
Jackson	September	16	39.0	6
Krout	Little Brown Hands	5	50.0	163
Larcom	Rivulet, The	4	80.0	297
Longfellow	Children's Hour, The	20	45.5	2
Longfellow	Hiawatha's Sailing	11	68.7	104
Mackay	Miller of the Dee, The	3	50.0	270
Moore	Visit from St. Nicholas, A	14	53.8	40
Nesbitt	Your Flag and My Flag	3	50.0	271
Rands	Great, Wide, Beautiful, Wonderful World	17	53.1	23
Riley	Little Orphant Annie	7	63.6	155
Rossetti	Swallow, The	2	40.0	308
Shaw	Columbia, the Gem of the Ocean	3	60.0	314
Sherman	Clouds	5	55.5	187
Sherman	Real Santa Claus, A	4	57.1	237
Sherman	Wizard Frost	4	50.0	212
Southey	Little Ladybird, The	2	40.0	316
Stedman	What the Winds Bring	6	66.7	188

TABLE VI—(*Continued*)

Author	Title	Frequency in Grade III	Per Cent of Total Frequency	Code No.
Taylor	Violet, The	8	66.7	147
Tennyson	Owl, The	7	63.6	157
Tennyson	Sweet and Low	14	42.2	19
Tennyson	Throstle, The	3	37.5	216
Thaxter	Little Gustava	3	37.5	217
Thaxter	March	4	36.4	158
Thomas	Talking in Their Sleep	4	44.4	190
Vandegrift	Sand Man, The	8	88.9	191
Van Dyke	Song Sparrow, The	2	40.0	326
Wordsworth	Kitten and the Falling Leaves, The	5	71.4	248
Wordsworth	To a Butterfly	7	58.3	148

TABLE VII—LIST OF POEMS FOR MEMORIZING IN GRADE IV

Author	Title	Frequency in Grade IV	Per Cent of Total Frequency	Code No.
Anonymous	Laughing Chorus, A	2	33.3	249
Browning	Songs from Pippa Passes	4	11.4	10
Bryant	Robert of Lincoln	11	44.0	43
Coleridge	He Prayeth Best	6	46.2	126
Deland	While Shepherds Watched Their Flocks by Night	2	40.0	288
Field	Night Wind, The	17	54.8	24
Gould	Frost, The	7	36.8	80
Hemans	Casabinanca	2	33.3	258
Hood	I Remember, I Remember	3	42.9	227
Hopkinson	Hail Columbia	2	18.2	152
Jackson	October's Bright Blue Weather	14	40.0	12
Kingsley	Farewell, A	5	50.0	162
Larcom	Calling the Violet	3	60.0	296
Longfellow	Arrow and the Song, The	11	31.4	13
Longfellow	Bell of Atri, The	4	50.0	203
Longfellow	Daybreak	6	37.5	103
Longfellow	Hiawatha's Friends	2	40.0	298
Longfellow	Rain in Summer	6	46.2	131
Longfellow	Village Blacksmith, The	20	47.6	4
Longfellow	Wreck of the Hesperus, The	4	50.0	205
Lowell	First Snowfall, The	11	32.4	15
Lowell	Fountain, The	6	50.0	141
MacDonald	Wind and the Moon, The	6	50.0	142
Payne	Home, Sweet Home	6	30.0	76
Procter, B. W	Sea, The	5	35.7	116
Riley	Brook Song, The	7	87.5	208
Riley	Sudden Shower, A	5	41.7	143
Tate	Christmas	3	30.0	173
Tennyson	Brook, The	7	28.0	46
Tennyson	Winter (Window, The)	2	28.6	244
Thaxter	Sandpiper, The	19	67.9	31
Whittier	Three Bells, The	3	60.0	327
Wordsworth	Lucy Gray	3	42.9	247
Wordsworth	Pet Lamb, The	2	40.0	329
Wordsworth	We are Seven	2	33.3	279
Wordsworth	Written in March	4	40.0	175

TABLE VIII—LIST OF POEMS FOR MEMORIZING IN GRADE V

Author	Title	Frequency in Grade V	Per Cent of Total Frequency	Code No.
Bennett	Flag Goes By, The	10	30.3	16
Bryant	Gladness of Nature, The	8	61.5	123
Bryant	Planting of the Apple Tree, The	10	52.6	78
Carlyle	Today	13	56.5	52
Cary, P.	Leak in the Dike, The	2	33.3	255
Emerson	Fable	11	52.4	67
Emerson	Forbearance	2	40.0	291
Hemans	Landing of the Pilgrims, The	13	39.4	17
Higginson	Four Leaf Clovers	3	37.5	199
Hogg	Skylark, The	0	00.0	259
Jackson	Down to Sleep	10	50.0	75
Keats	On the Grasshopper and Cricket	3	60.0	294
Key	Star Spangled Banner	11	28.9	8
Longfellow	Day is Done, Thy	14	56.0	44
Longfellow	Excelsior	8	66.7	140
Longfellow	Old Clock on the Stairs, The	5	62.5	204
Longfellow	Paul Revere's Ride	9	37.5	49
Longfellow	Ship of State, The	6	37.5	105
Lowell	Aladdin	4	57.1	230
Lowell	Heritage, The	3	37.5	206
Milton	Song on a May Morning	3	60.0	303
Morris	Woodman, Spare That Tree	9	56.3	106
Riley	Song, A	3	30.0	168
Riley	Knee Deep in June	3	33.3	185
Saxe	Blind Men and the Elephant, The	2	40.0	311
Shakespeare	Under the Greenwood Tree (As You Like It)	6	42.9	117
Southey	Inchcape Rock, The	4	50.0	213
Thaxter	Piccola	2	40.0	325
Whittier	Barbara Frietchie	9	69.2	138
Whittier	Barefoot Boy, The	15	45.5	20
Whittier	Corn Song, The	11	42.3	42
Whittier	In School Days	8	44.4	88
Wordsworth	My Heart Leaps Up When I Behold	2	25.0	218

TABLE IX—LIST OF POEMS FOR MEMORIZING IN GRADE VI

Author	Title	Frequency in Grade VI	Per Cent of Total Frequency	Code No.
Aldrich	Before the Rain	11	78.6	113
Blake	Tiger, The	2	33.3	251
Brooks	Christmas Everywhere	4	36.3	149
Browning	How They Brought the Good News	4	50.0	194
Browning	Pied Piper of Hamelin, The	2	40.0	282

TABLE IX—(*Continued*)

Author	Title	Frequency in Grade VI	Per Cent of Total Frequency	Code No.
Bryant	March	7	46.7	107
Bryant	To the Fringed Gentian	3	23.1	124
Bryant	Yellow Violet, The	2	40.0	283
Campbell	Lord Ullin's Daughter	5	83.3	254
Cary, A.	Nobility	10	58.8	90
Cary, A.	Order for a Picture, An	2	28.6	221
Drake	American Flag, The	8	28.6	29
Emerson	Concord Hymn	8	29.6	32
Finch	Blue and the Gray, The	9	52.9	92
Holland	Christmas Carol, A	3	42.9	226
Holmes	Old Ironsides	14	51.9	33
Hunt	Abou Ben Adhem	15	42.9	11
Kingsley	Three Fishers, The	2	33.3	262
Lang	Scythe Song	4	36.6	264
Longfellow	Builders, The	14	60.9	57
Longfellow	Psalm of Life, A	13	59.1	63
Longfellow	Sandalphon	3	60.0	300
Lowell	To the Dandelion	5	50.0	165
Mackay	Tubal Cain	2	28.6	232
Moore	Minstrel Boy, The	3	60.0	304
Perry	Coming of Spring, The	4	40.0	167
Procter, A.	One by One	3	50.0	272
Read	Sheridan's Ride	6	54.5	154
Riley	Old Glory	5	29.4	95
Scott	Lochinvar	3	37.5	209
Shakespeare	Hark, Hark! the Lark (Cymbeline)	1	11.1	186
Shakespeare	Orpheus With His Lute (King Henry the Eighth)	5	71.4	236
Sherman	May	3	50.0	274
Tennyson	Break, Break, Break	7	53.8	136
Tennyson	Bugle Song	11	42.3	41
Tennyson	Charge of the Light Brigade	10	45.5	66
Tennyson	Eagle, The	1	20.0	323
Tennyson	Ring Out, Wild Bells	4	40.0	174
Wolfe	Burial of Sir John Moore	5	55.6	192
Wordsworth	Fidelity	3	60.0	328
Wordsworth	I Wandered Lonely as a Cloud	10	30.3	21
Wordsworth	To a Skylark	5	55.6	193

TABLE X—LIST OF POEMS FOR MEMORIZING IN GRADE VII

Author	Title	Frequency in Grade VII	Per Cent of Total Frequency	Code No.
Addison	Spacious Firmament on High, The	5	38.5	122
Alexander	Burial of Moses, The	2	40.0	280
Browning	Home Thoughts from Abroad	4	57.1	220
Browning	Incident of the French Camp, An	4	36.4	150

TABLE X—(*Continued*)

Author	Title	Frequency in Grade VII	Per Cent of Total Frequency	Code No.
Bryant	Death of the Flowers, The	2	22.2	177
Bryant	Song of Marion's Men	5	55.6	178
Bryant	To a Waterfowl	8	33.3	47
Burns	Bannockburn	3	50.0	252
Burns	My Heart's in the Highlands	3	60.0	284
Campbell	Hohenlinden	4	50.0	195
Emerson	Rhodora, The	2	33.3	256
Emerson	Snowstorm, The	5	38.5	127
Holland	Gradatim	7	30.4	56
Holmes	Union and Liberty	5	50.0	160
Longfellow	Building of the Ship, The	7	63.6	153
Longfellow	Courtship of Miles Standish, The	5	83.3	266
Longfellow	Evangeline	5	62.5	202
Longfellow	King Robert of Sicily	2	40.0	299
Longfellow	Ladder of St. Augustine, The	2	33.3	267
Longfellow	Rainy Day, The	4	40.0	164
Lowell	Fatherland, The	4	30.8	133
Lowell	Finding of the Lyre	5	71.4	231
Lowell	Yussouf	2	25.0	207
Macaulay	Horatius	4	44.4	184
Miller	Columbus	11	44.0	45
Riley	Life Lesson, A	3	42.9	233
Shakespeare	Good Name (Othello)	4	50.0	210
Stoddard	Abraham Lincoln	2	40.0	322
Taylor	Song of the Camp, The	3	42.9	240
Tennyson	Crossing the Bar	5	38.5	137
Tennyson	In Memoriam	3	42.9	241
Tennyson	Sir Galahad	4	57.1	243
Whittier	Huskers, The	3	42.9	246
Wilder	Stand by The Flag	3	50.0	278

TABLE XI—LIST OF POEMS FOR MEMORIZING IN GRADE VIII

Author	Title	Frequency in Grade VIII	Per Cent of Total Frequency	Code No.
Bryant	Thanatopsis	10	71.4	114
Burns	For A' That and A' That	8	53.3	108
Byron	Destruction of Sennacherib	3	60.0	285
Byron	Eve of Waterloo, The	5	83.3	253
Emerson	Each and All	3	60.0	290
Gray	Elegy Written in a Country Churchyard	8	88.9	181
Holmes	Chambered Nautilus, The	17	56.7	26
Holmes	Last Leaf, The	7	46.7	109
Howe	Battle Hymn of the Republic	10	37.0	34
Jones	What Constitutes a State?	4	66.7	261
Kipling	If	3	50.0	263
Kipling	L'Envoi	5	71.4	228
Kipling	Recessional	19	79.2	48
Lanier	Song of the Chattahoochee	4	57.1	229

TABLE VIII—(*Continued*)

Author	Title	Frequency in Grade V	Per Cent of Total Frequency	Code No.
Lowell	Day in June, A	7	53.8	132
Lowell	Ode Recited at the Harvard Commemoration	5	100.0	301
Lowell	Stanzas on Freedom	5	83.3	268
Lowell	Vision of Sir Launfal, The	11	52.4	69
Milton	On His Blindness	4	80.0	302
Newman	Lead Kindly Light	4	40.0	166
Pierpont	Warren's Address	7	36.8	84
Scott	Breathes There a Man	15	37.5	7
Scott	Lady of the Lake	4	80.0	312
Shakespeare	Mark Anthony's Address (Julius Caesar)	7	100.0	235
Shakespeare	Mercy Speech (Merchant of Venice)	17	70.8	50
Shakespeare	Polonius to Laertes (Hamlet)	11	91.7	144
Shelley	Cloud, The	5	35.7	118
Shelley	Skylark, The	6	60.0	169
Sill	Opportunity	10	66.7	111
Tennyson	Flower in the Crannied Wall	3	37.5	215
Tennyson	King Arthur	3	42.9	242
Whitman	Oh Captain! My Captain!	17	73.9	60
Whittier	Snowbound: A Winter Idyll	8	42.1	85

The number of poems in each grade list as shown in Tables IV to XI is as follows:

Grade	Number of Poems
I	46
II	60
III	45
IV	36
V	33
VI	42
VII	35
VIII	32
Total	329

Why More Poems are Given for Lower Than for Upper Grades—It is evident that there are more poems for the lower than for the upper grades. This difference may be due in part to the greater ease of choosing appropriate poems for the lower grades. Several of the authors who contributed the greatest number of poems to the Selected List are authors whose poems have been written largely for young children. For example, there are 23 poems by Stevenson all of which are suggested for Grades I and II. Similarly, seven of the 10 poems by Field are suggested for Grades I and II. Nine of the 13 poems by Sherman are for the first two grades, and 5 of Rossetti's 6 poems are for the same grades.

Another reason why our material led to the selection of a greater number of poems for the lower grades was that, in general, the courses of study presented relatively shorter lists of poems in the upper grades. Indeed, several courses of study did not contain any lists at all for grades above the fourth.

Again, there is an evident tendency on the part of makers of courses of study to assign poems to the lowest grade to which they are thought to be appropriate. For example, as has been pointed out, patriotic poems are often assigned for very young children to learn. Other poems are likewise required in the primary grades which better judgment might place in higher grades. Tables IV to XI reveal numerous examples of this. Tennyson's "Sweet and Low" is assigned to the third grade. Yet the *thought* of this lyric, apart from its music, is no doubt quite beyond the understanding of third grade children. Likewise, Longfellow's "Children's Hour" is assigned to the third grade. Its vocabulary, its allusion, its atmosphere of New England culture—these are by no means familiar to the majority of third grade children. It is true the poem is *about* children; but a poem about children may be very far from a child's poem. At any rate, the tendency to grade poems down is much more marked in these courses of study than the tendency to grade them up; and this, as well as the conditions above mentioned, accounts for the fact that although the literature appropriate to the understanding of upper grade children is no doubt richer than that which appeals to lower grade children, the poems selected for the latter to memorize considerably outnumber those selected for the former.

VI—VERIFYING THE GRADING OF THE SELECTED LIST

Comparison with Atherton's Grading—Atherton likewise grouped some of his selections into grade lists according to the grades in which they were most frequently noted. For this purpose he used only the selections which were mentioned at least ten times. These grade lists contained 117 of his original 382 titles; and of the 117 titles, 111 also appear in our graded lists (Tables IV to XI). The remaining six titles not on our lists, are: (1) MacDonald's "Baby", (2) Stedman's "What the Winds Bring," (3) Scott's "Love of Country", (4) Lincoln's "Gettysburg Address", (5) "The Twenty-third Psalm", (6) "Mother Goose." Only two of these are really additions to the Selected List. Our list does not contain prose selections, psalms, or mother goose rhymes. Moreover, Scott's "Love of Country" is the same selection as "Breathes There a Man", which appears in Atherton's Grade Lists and also in ours.

Table XII shows the correspondence between the grade assignments of the 111 titles common to the two investigations. The correlation between the grade assignments of the two lists is +0.84. Eighty-one of the 111 poems or 73 percent were placed in the same grade by both studies. The table shows a slight though consistent tendency for a poem to be introduced into a lower grade by our lists than by Atherton's lists. In other words, our lists are a little more exacting.

The correlation ratios which express more generally—and especially when the regression lines are not rectilinear—the relationship between two series of measures, were computed with the result that η_x equals 0.945 and η_y equals 0.943. These correlation ratios differ somewhat from the correlation coefficient although the difference is not very great. The difference, however, suggests that the relationship between the grading by the Atherton list and the grading by our list may not

TABLE XII—CORRESPONDENCE IN THE GRADING OF THE 111 POEMS COMMON TO THE SELECTED LIST AND ATHERTON'S GRADE LIST

	Grading According to Selected List								Total
	I	II	III	IV	V	VI	VII	VIII	
Grading According to Atherton I	8	2							10
II	1	15							16
III		1	11	1					13
IV		2	3	9	1				15
V				4	14	1			19
VI				1	2	9	1	1	14
VII				1	2	3	4	2	12
VIII					1		1	10	12
Total	9	20	14	16	20	13	6	13	111

$r=0.836$
$\eta_x=0.945$
$\eta_y=0.943$

be exactly rectilinear—in other words, that equal differences in the scale of grading according to Atherton do not correspond troughout the range to equal differences in grading according to our list. The point, however, is rather too remote from practical utility to be pressed to its statistical conclusion. On its face Table XII shows a close and regular relationship between the two sources of grading.

Comparison with Grading in Readers—An examination of the grading of the poetry appearing in 102 of the most popular elementary school readers was also made; and the results were compared with those shown in the Selected List. The readers ranged from those for the first grade to those for the eighth grade.* These readers contained 272 poems which were also in the Selected List. A grading of the poems as they appeared in the readers was made by the same method used in grading the Selected List. The frequency of mention ranged from 1 to 10. Two poems, "Abou Ben Adhem," by Leigh Hunt, and "The Wind" by Rossetti, occurred in 10 of the readers. Since most of the poems appeared but relatively few times in the readers, the grading according to them is necessarily less reliable than is the grading of the Selected List.

Table XIII shows the relationship between our grading of the poems and that of the readers. A greater scattering in grade assignments is evident in Table XIII than in Table XII (which shows the relationship between Atherton's grading and the grading of the Selected List). In the first four grades, courses of study tend to assign poems to lower grades than do readers. In other words, school practice places more difficult poems in the hands of young children for memorizing than merely for reading. This is not difficult to explain. Obviously, children can memorize poetry long before they can read it. On the other hand, in the upper grades courses of study are less exacting than readers. It appears to be assumed that, with reference to poetry, ability to read has not only caught up with ability to memorize but has actually become superior to it.

The correlation coefficient by the product moment formula is +0.85 which is slightly larger than the correlation between Atherton's grading and the grading of the Selected List. Yet only 95 poems, or 34.9 percent of those common to the readers and the Selected List, are assigned to the same grades, as opposed to 73 percent for Atherton's Grade List and the Selected List.

---

*See appendix for list of readers.

TABLE XIII—CORRESPONDENCE IN THE GRADING OF THE 272 POEMS COMMON TO THE SELECTED LIST AND TO THE ELEMENTARY READERS

		Grading According to Selected List								Total
		I	II	III	IV	V	VI	VII	VIII	
Grading According to Readers	I	11								11
	II	17	14	2	3					36
	III	5	28	13			1			47
	IV	3	5	13	15	5	2	1		44
	V			6	12	16	14	4	5	57
	VI				2	5	5	13	3	28
	VII			1		4	7	10	8	30
	VIII			1			4	3	11	19
Total		36	47	36	32	30	33	31	27	272

$r=0.850$

$\eta_x=0.853$

$\eta_y=0.867$

This again raises the question of the sufficiency of the correlation coefficient as a measure of relationship. A casual inspection of Table XIII may lead one to believe that the relationship between the two methods of grading is not rectilinear. This suspicion, however, is not borne out by further statistical analysis. The correlation ratios are: η_x equals 0.853 and η_y equals 0.867. These are so nearly equal to the correlation coefficient that the rectilinearity of the regression lines appears to be established. The test for a rectilinear relationship rests upon a comparison of each Eta-value with the r-value. If the values are equal, the regression lines are straight. Here the differencts between the Eta-values and the r-value are so small that they are quite within the variability which may be expected by chance.

On the whole, it will appear evident, we think, that both Atherton's grading and the grading by readers justify us in the belief that our grading is reasonably accurate.

*V*II—PREFERRED LISTS OF POEMS FOR MEMORY WORK

The 81 poems, placed in the same grades by both Atherton's Grade Lists and by ours, were examined with reference to the following criteria: (1) Do they have a high frequency in the grade to which they are assigned; and (2) do they have a high total frequency. Practically every one of these poems met the criteria. We have therefore made up for each grade a "Preferred List" consisting of poems agreed upon as to grading by Atherton and ourselves, together with a few other poems which according to our lists also met the criteria above mentioned. The poems in roman type are those which were placed in the given grade by both Atherton's study and by ours. The poems in italics are those which according to our lists are appropriate, but which were not so graded by Atherton. Table XIV shows the total frequency of mention which each poem received according to our investigation and also its percent of frequency in the grade to which it was assigned. Patriotic poems were purposely omitted, because their grading is not thought to have the same meaning that grading has in the case of other poems.

TABLE XIV—PREFERRED LISTS OF POEMS FOR MEMORY WORK

Grade I

Author	Title	Total Frequency	Percent of Frequency in Grade I
1. Alexander, Mrs. C. F.	All Things Bright and Beautiful	20	50.0
2. Brown, Kate L.	Little Plant, The	22	95.4
3. From the German	Sleep, Baby, Sleep	17	88.2
4. Stevenson, R. L.	Autumn Fires	14	57.1
5. Stevenson, R. L.	Rain	22	90.9
6. Stevenson, R. L.	Swing, The	32	71.9
7. Taylor, Jane	Twinkle, Twinkle, Little Star	23	91.3
8. Tennyson, Lord	Bird and the Baby, The	31	93.6
9. *Cooper, Geo.*	*Come Little Leaves*	18	55.5
10. *Field, Eugene*	*Why Do Bells for Christmas Ring?*	29	48.3
11. *Rossetti, Christina*	*Wind, The*	23	86.9
12. *Stevenson, R. L.*	*My Shadow*	42	50.0

TABLE XIV—(*Continued*)

Grade II

Author	Title	Total Frequency	Percent of Frequency in Grade II
1. Allingham, Wm.	Fairies, The	17	41.2
2. Cary, Phoebe	Suppose	20	45.0
3. Child, Lydia M.	Thanksgiving Day	29	41.4
4. Coleridge, S. T.	Answer to a Child's Question	12	50.0
5. Field, Eugene	Dutch Lullaby	43	51.2
6. Field, Eugene	Rockabye Lady, The	16	43.8
7. Houghton, Lord	Lady Moon	19	47.4
8. Ingelow, Jean	Seven Times One	33	54.5
9. Kingsley, Chas.	Lost Doll, The	19	42.1
10. Larcom, Lucy	Brown Thrush, The	27	55.5
11. Lear, Edward	Owl and the Pussy Cat, The	14	35.7
12. Sherman, F. D.	Daisies	27	55.5
13. Sherman, F. D.	Four Winds, The	15	46.7
14. Stevenson, R. L.	Land of Story Books, The	27	44.4
15. *Field, Eugene*	*Duel, The*	17	76.5
16. *Longfellow, H. W.*	*Hiawatha's Childhood*	27	59.3
17. *Miller, Emily*	*Blue Bird, The*	28	46.4
18. *Stevenson, R. L.*	*Wind, The*	38	42.1
19. *Stevenson, R. L.*	*Windy Nights*	21	47.6

Grade III

Author	Title	Total Frequency	Percent of Frequency in Grade III
1. Aldrich, T. B.	Marjorie's Almanac	18	66.7
2. Allingham, Wm.	Wishing	20	60.0
3. Bjornson, Bjornsterne	Tree, The	20	55.5
4. Brooks, Phillips	O Little Town of Bethlehem	22	36.4
5. Cary, Alice	November	23	39.1
6. Field, Eugene	Norse Lullaby	19	57.9
7. Hogg, James	Boy's Song	23	47.8
8. Krout, Mary H.	Little Brown Hands	10	50.0
9. Longfellow, H. W.	Hiawatha's Sailing	16	68.7
10. Moore, C. C.	Visit from St. Nicholas, A	26	53.8
11. Tennyson, Lord	Owl, The	11	63.6
12. *Jackson, Helen H.*	*September*	41	39.0
13. *Longfellow, H. W.*	*Children's Hour, The*	44	45.5
14. *Rands, W. B.*	*Great, Wide, Beautiful, Wonderful World*	32	53.1

TABLE XIV—(*Continued*)

Grade IV

Author	Title	Total Freq-uency	Percent of Frequency in Grade IV
1. Field, Eugene	Night Wind, The	31	48.4
2. Gould, Hannah F.	Frost, The	19	36.8
3. Jackson, Helen H.	October's Bright Blue Weather	35	45.7
4. Longfellow, H. W.	Village Blacksmith, The	42	47.6
5. Lowell, J. R.	Fountain, The	12	50.0
6. Tate, Nahum	Christmas	10	30.0
7. Tennyson, Lord	Brook, The	25	28.0
8. Thaxter, Celia	Sandpiper, The	28	67.9
9. Wordsworth, Wm.	Lucy Gray	7	42.9
10. *Bryant, W. C.*	*Robert of Lincoln*	25	44.0
11. *Longfellow, H. W.*	*Arrow and the Song, The*	35	31.4
12. *Lowell, J. R.*	*First Snowfall, The*	34	32.4
13. *Riley, J. W.*	*Brook Song, The*	8	87.5

Grade V

Author	Title	Total Freq-uency	Percent of Frequency in Grade V
1. Bryant, W. C.	Gladness of Nature, The	13	61.5
2. Bryant, W. C.	Planting of the Apple Tree, The	19	52.6
3. Carlyle, Thomas	Today	23	56.5
4. Hemans, Felicia	Landing of the Pilgrims, The	33	39.4
5. Jackson, Helen H.	Down to Sleep	20	50.0
6. Longfellow, H. W.	Excelsior	12	66.7
7. Longfellow, H. W.	Paul Revere's Ride	24	37.5
8. Longfellow, H. W.	Day is Done, The	25	56.0
9. Longfellow, H. W.	Old Clock on the Stairs, The	8	62.5
10. Morris, G. P.	Woodman, Spare That Tree	16	56.3
11. Shakespeare, Wm.	Under the Greenwood Tree	14	42.9
12. Whittier, J. G.	In School Days	17	44.4
13. Whittier, J. G.	Barefoot Boy, The	33	42.4
14. Whittier, J. G.	Corn Song, The	26	42.3
15. *Emerson, R. W.*	*Fable*	21	52.4
16. *Whittier, J. G.*	*Barbara Frietchie*	13	69.2

TABLE XIV—(*Continued*)

Grade VI

Author	Title	Total Frequency	Percent of Frequency in Grade VI
1. Aldrich, T. B.	Before the Rain	14	78.6
2. Cary, Alice	Nobility	17	58.8
3. Drake, J. R.	American Flag, The	28	28.6
4. Holmes, O. W.	Old Ironsides	27	51.9
5. Hunt, Leigh	Abou Ben Adhem	35	42.9
6. Longfellow, H. W.	Builders, The	23	60.9
7. Tennyson, Lord	Charge of the Light Brigade	22	45.5
8. Wolfe, Chas.	Burial of Sir John Moore	9	55.6
9. Wordsworth, Wm.	I Wandered Lonely as a Cloud	33	30.3
10. *Longfellow, H. W.*	*Psalm of Life, A*	22	59.1
11. *Tennyson, Lord*	*Break, Break, Break*	13	53.8
12. *Tennyson, Lord*	*Bugle Song*	26	42.3

Grade VII

Author	Title	Total Frequency	Percent of Frequency in Grade VII
1. Bryant, W. C.	To a Waterfowl	24	33.3
2. Emerson, R. W.	Snowstorm, The	13	38.5
3. Holland, J. G.	Gradatim	23	30.4
4. Miller, Joaquin	Columbus	25	44.0
5. *Browning, R.*	*Home Thoughts from Abroad*	7	57.1
6. *Longfellow, H. W.*	*Building of the Ship, The*	11	63.6
7. *Lowell, J. R.*	*Fatherland, The*	13	30.8
8. *Lowell, J. R.*	*Yussouf*	8	25.0
9. *Macaulay, T. B.*	*Horatius*	9	44.4
10. *Taylor, Bayard*	*Song of the Camp, The*	7	42.9
11. *Tennyson, Lord*	*Sir Galahad*	7	57.1

Grade VIII

Author	Title	Total Frequency	Percent of Frequency in Grade VIII
1. Bryant, W. C.	Thanatopsis	14	71.4
2. Holmes, O. W.	Last Leaf, The	15	46.7
3. Kipling, Rudyard	Recessional	24	79.2
4. Pierpont, John	Warren's Address	19	36.8
5. Shakespeare, Wm.	Mercy Speech	24	70.8
6. Shelley, P. B.	Cloud, The	14	35.7
7. Sill, E. R.	Opportunity	15	66.7
8. Whitman, Walt	O Captain! My Captain!	23	73.9
9. Whittier, J. G.	Snowbound	19	42.1
10. *Burns, Robert*	*For A' That and A' That*	15	53.3
11. *Gray, Thomas*	*Elegy Written in a Country Churchyard*	9	88.9
12. *Holmes, O. W.*	*Chambered Nautilus, The*	30	56.7
13. *Lowell, J. R.*	*Day In June, A*	13	53.8
14. *Scott, Sir W.*	*Breathes There a Man*	40	37.5
15. *Shakespeare, Wm.*	*Polonius to Laertes*	12	91.7

These Preferred Lists probably constitute the best short lists which an investigation of present courses of study can yield. Such lists, it is to be understood, are based on agreement between courses of study. If there is any constant bias—any convention which has grown up whereby one writer of a course imitates others—these results will likewise be biased in the same direction. If, for example, there is a tendency to assign to elementary school children poems which they cannot understand or appreciate, this tendency will be evident in the resulting lists. A useful criticism of these lists would be afforded by reading the poems to children and obtaining their preferences. At present, however, it is important to know what the current requirements are. Many of them are unquestionably good; and we feel sure that our lists are more reliable and more useful than those contained in any single published course of study.

VIII—The Finding List

Table XV was devised primarily as a "finding list." The first column contains in alphabetical order the names of authors who contributed to the Selected List, together with the titles of their poems. It also shows the nationality and dates of birth and death of the authors so far as it has been possible to obtain these facts. The number in parenthesis following the title is the code number of the poem and may be used to locate it in Tables II and III. The second column gives the first line of each poem. There is sometimes a question as to the identity of a given poem or selection—a question which the first line will usually clear up. Or a poem may be better known by its opening lines than by its title. For these reasons the poems have been examined and the first lines have been quoted. The column headed "Grade" indicates the grade to which each poem has been assigned, and the last column headed "Reference" gives the code numbers of compilations of poetry which contain the various poems. The titles of these compilations and their code numbers may be found in the bibliography in Part III.

TABLE XV—THE SELECTED LIST ORGANIZED AS A FINDING LIST

Author and Title	First Line	Grade	Reference
1. Addison, Joseph (English 1672-1719)			
(a) Spacious Firmament on High, The (122)	The spacious firmament on high	VII	25, 26, 30
2. Aldrich, Thomas Bailey (American 1836-1907)			
(a) Before the Rain (113)	We knew it would rain, for all the morn	VI	4, 30
(b) Marjorie's Almanac (86)	Robins in the tree top	III	1, 32
3. Alexander, Cecil (Mrs.) (Irish 1818-1895)			
(a) All Things Bright and Beautiful (71)	All things bright and beautiful	I	9, 32
(b) Burial of Moses, The (280)	By Nebo's lonely mountain	VII	25, 30
4. Allingham, Wm. (Irish 1828-1889)			
(a) Fairies, The (89)	Up the airy mountain	II	9, 25, 26 30, 32
(b) Robin Redbreast (98)	Good-bye, good-bye to summer!	II	9, 25, 26 32
(c) Wishing (72)	Ring-ting! I wish I were a Primrose	III	25, 26, 32
5. Allison, Joy (Cragin, Mrs. Mary)			
(a) Which Loved Best (99)	"I love you mother," said little John	II	15,
6. Anonymous			
(a) Laughing Chorus, A (249)	Oh, such a commotion under the ground	IV	12
(b) Secret, The (281)	We have a secret just we three	I	12
(c) Sleep, Baby, Sleep, (93)*	Sleep, baby, sleep	I	31
7. Bangs, John Kendrick (American 1862-)			
(a) Little Elf, The (250)	I met a little elf man once,	II	24, 25, 26 32
8. Bates, Clara Doty (American 1838-1895)			
(a) Who Likes the Rain? (219)	"I," said the duck, "I call it fun	I	31
9. Bennett, Henry Holcomb (American 1863-)			
(a) Flag Goes By, The (16)	Hats off! Along the street there comes	V	1, 24, 25 26, 30

*From the German

TABLE XV—(*Continued*)

Author and Title	First Line	Grade	Reference
10. Bjornson, Bjornsterne (Norwegian 1832-1910)			
(a) Tree, The (73)	I love thee when thy swelling buds appear	III	25, 26, 30 32
11. Blake, Wm. (English 1757-1827)			
(a) Lamb, The (159)	Little lamb, who made thee?	I	9, 25, 26 32
(b) Tiger, The (251)	Tiger, tiger burning bright	VI	9, 25, 26 30
12. Brooks, Phillips (American 1835-1893)			
(a) Christmas Everywhere (149)	Everywhere, everywhere, Christmas tonight!	VI	1
(b) O Little Town of Bethlehem (61)	O little town of Bethlehem	III	24, 25, 30
13. Brown, Kate Louise (American 1857-)			
(a) Little Plant, The (62)	In the heart of a seed	I	12
14. Browning, Elizabeth Barrett (Mrs.) (English 1809-1861)			
(a) Child's Thought of God, A (176)	They say that God lives very high!	III	32
15. Browning, Robert (English 1812-1889)			
(a) Home Thoughts from Abroad (220)	Oh, to be in England	VII	25, 26
(b) How They Brought the Good News (194)	I sprang to the stirrup, and Joris and he:	VI	9, 25, 26 30
(c) Incident of the French Camp, An (150)	You know we French stormed Ratisbon:	VII	9, 25, 26 30
(d) Pied Piper of Hamelin, The (282)	Hamelin town's in Brunswick	VI	9, 25, 26 30
(e) Songs from Pippa Passes (10)	The year's at the spring	IV	9, 25, 26
16. Bryant, William Cullen (American 1794-1878)			
(a) Death of the Flowers, The (177)	The melancholy days are come, the saddest of the year	VII	24, 25, 30
(b) Gladness of Nature, The (123)	Is this a time to be cloudy and sad,	V	25, 26
(c) March (107)	The stormy March is come at last	VI	30
(d) Planting of the Apple Tree, The (78)	Come, let us plant the apple tree	V	24, 25, 26 30
(e) Robert of Lincoln (43)	Merrily swinging on brier and weed	IV	9, 25, 26 32
(f) Song of Marion's Men (178)	Our band is few, but true and tried	VII	25, 26
(g) Thanatopsis (114)	To him who, in the love of Nature holds,	VIII	1, 24, 25 26

TABLE XV—(*Continued*)

Author and Title	First Line	Grade	Reference
16. Bryant, William Cullen (Cont.)			
(h) To a Waterfowl (47)	Whither 'midst falling dew,	VII	1, 24, 25 26, 30
(i) To the Fringed Gentian (124)	Thou blossom bright with autumn dew	VI	24, 25, 31
(j) Yellow Violet, The (283)	When beechen buds begin to swell,	VI	5
17. Bunner, Henry Cuyler (American 1855-1896)			
(a) One, Two, Three (51)	It was an old, old, old, old lady	II	9, 32
18. Burns, Robert (Scotch 1759-1796)			
(a) Bannockburn (252)	Scots, wha hae wi' Wallace bled	VII	6, 30
(b) For A' That and A' That (Is There For Honest Poverty) (108)	Is there, for honest poverty	VIII	1, 25, 26
(c) My Heart's in the Highlands (284)	My heart's in the highlands, my heart is not here:	VII	9, 25, 30
19. Byron, Lord (George Gordon) (English 1788-1824)			
(a) Destruction of Sennacherib (285)	The Assyrian came down like the wolf on the fold,	VIII	25
(b) Eve of Waterloo, The (253)	There was a sound of revelry by night	VIII	9, 25, 30
20. Campbell, Thomas (English 1777-1844)			
(a) Hohenlinden (195)	On Linden when the sun was low,	VII	4, 25, 30
(b) Lord Ullin's Daughter (254)	A chieftain to the Highlands bound	VI	9, 25, 26 30
21. Carlyle, Thomas (English 1795-1881)			
(a) Today (52)	So here hath been dawning	V	9, 25, 26 30
22. Cary, Alice (American 1820-1871)			
(a) Nobility (90)	True worth is in being, not seeming,	VI	1, 8
(b) November (53)	The leaves are fading and falling,	III	8
(c) Order for a Picture, An (221)	Oh, good painter, tell me true	VI	8
23. Cary, Phoebe (American 1824-1871			
(a) Don't Give Up (125)	If you've tried and have not won,	III	8
(b) Leak in the Dike, The (255)	The good dame looked from her cottage	V	8

TABLE XV—(*Continued*)

Author and Title	First Line	Grade	Reference
23. Cary, Phoebe (Cont.)			
(c) Obedience (286)	If you're told to do a thing	II	8
(d) Suppose (74)	Suppose, my little lady	II	8
(e) They Didn't Think (222)	Once a trap was baited,	I	8
24. Child, Lydia Maria (American 1802-1880)			
(a) Thanksgiving Day (27)	Over the river and through the wood,	II	32
(b) Who Stole the Bird's Nest? (100)	"To-whit! to-whit! to-whee!"	II	9, 32
25. Coleridge, Samuel Taylor (English 1772-1834			
(a) Answer to a Child's Question (139)	Do you ask what the birds say? The Sparrow, the Dove,	II	25, 26
(b) He Prayeth Best (126)	He prayeth best, who loveth best,	IV	25
26. Cone, Helen Gray (American 1859-)			
(a) Dandelions, The (179)	Upon a showery night and still,	III	25, 32
27. Coolidge, Susan (Woolsey, Sarah) (American 1845-1905)			
(a) How the Leaves Came Down (54)	I'll tell you how the leaves came down	II	25, 26, 32
28. Cooper, George (American 1820-1876)			
(a) Come, Little Leaves (87)	"Come little leaves," said the wind one day,	I	31
(b) October's Party (223)	October gave a party	II	25, 26
(c) What Robin Told (225)	How do robins build their nest?	II	12
(d) Wonderful Weaver, The (224)	There's a wonderful weaver	III	12
29. Cowper, William (English 1731-1800)			
(a) The Nightingale and the Glowworm (287)	A nightingale that all day long	III	6, 25, 32
30. Deland, Mary Margaret (American 1857-)			
(a) While Shepherds Watched Their Flocks by Night (288)	Like small curled feathers, white and soft	IV	26, 30
31. Dodge, Mary Mapes (Mrs.) (American 1838-1905)			
(a) Snowflakes (289)	Whenever a snowflake leaves the sky,	II	24, 26, 32
32. Drake, Joseph Rodman (American 1795-1820)			
(a) American Flag, The (29)	When Freedom from her mountain height	VI	4, 24, 25 26, 30

TABLE XV—(*Continued*)

Author and Title	First Line	Grade	Reference
33. Edwards, Matilda Betham (English 1836-)			
(a) Child's Prayer, A (196)	God make my life a little light	III	9, 32.
34. Emerson, Ralph Waldo (American 1803-1882)			
(a) Concord Hymn (32)	By the rude bridge that arched the flood,	VI	1, 4, 24 25, 26, 30
(b) Each and All (290)	Little thinks, in the field, yon red-cloaked clown	VIII	1, 24, 25
(c) Fable (67)	The mountain and the squirrel,	V	25, 32
(d) Forbearance (291)	Hast thou named all the birds without a gun?	V	24, 25, 26 30
(e) Rhodora, The (256)	In May, when sea-winds pierced our solitudes,	VII	1, 24, 25 26, 30
(f) Snowstorm, The (127)	Announced by all the trumpets of the sky	VII	24, 30
(g) We Thank Thee (128)	For gainful hours of pain and loss	II	10
35. Field, Eugene (American 1850-1895)			
(a) Duel, The (91)	The gingham dog and the calico cat	II	1, 9, 11 25, 26
(b) Dutch Lullaby (Wynken, Blynken and Nod) (3)	Wynken, Blynken and Nod one night	II	6, 9, 11 24, 25, 26 32
(c) Japanese Lullaby (101)	Sleep, little pigeon, and fold your wings	II	11
(d) Little Boy Blue (39)	The little toy dog is covered with dust	I	1, 9, 11 24, 25
(e) Night Wind, The (24)	Have you ever heard the wind go "Yooooo?"	IV	11
(f) Norse Lullaby (79)	The sky is dark and the hills are white	III	11
(g) Rockabye Lady, The (102)	The Rockabye Lady from Hushabye Street	II	9, 11
(h) Seein' Things (197)	I ain't afeared uv snakes, or toads, or bugs, or worms, or mice	III	11
(i) Sugar Plum Tree, The (129)	Have you ever heard of the sugar plum tree?	I	9, 11, 25 26
(j) Why do Bells for Christmas Ring? (28)	Why do the bells for Christmas ring?	I	11
36. Fields, John T. (American 1817-1881)			
(a) Ballad of the Tempest, The (Captain's Daughter, The) (151)	We were crowded in the cabin	III	4, 6, 25 26
37. Finch, Francis Miles (American 1827-1907)			
(a) Blue and the Gray, The (92)	By the flow of inland river,	VI	1, 4, 24 25

TABLE XV—(*Continued*)

Author and Title	First Line	Grade	Reference
38. Follen, Eliza Lee (Cabot) (American 1787-1860)			
(a) New Moon, The (130)	Dear mother, how pretty	II	1, 31
39. Garabrant, Nellie			
(a) Dandelion (180)	There's a dandy little fellow	II	12
40. Gould, Hannah F. (American 1789-1865)			
(a) Frost, The (80)	The frost looked forth on a still, clear night	IV	6, 26, 32
41. Gray, Thomas (English 1716-1771)			
(a) Elegy Written in a Country Churchyard (181)	The curfew tolls the knell of parting day,	VIII	1, 4, 25 26, 30
42. Hale, Sarah Josepha (American 1790-1879)			
(a) Mary's Lamb (257)	Mary had a little lamb,	I	12
43. Hemans, Felicia (Mrs.) (English 1793-1835)			
(a) Casabianca (258)	The boy stood on the burning deck	IV	6, 9, 25 26
(b) Landing of the Pilgrims, The (17)	The breaking waves dashed high	V	6, 25, 26 30
44. Herford, Oliver (American 1863-)			
(a) Elf and the Dormouse, The (198)	Under a toadstool	II	9, 24, 25 26, 32
45. Higginson, Ella (American 1862-)			
(a) Four Leaf Clovers (199)	I know a place where the sun is like gold	V	20, 24, 25
46. Hogg, James (Scotch 1770-1835)			
(a) Boy's Song (55)	Where the pools are bright and deep	III	6, 9, 25 26, 32
(b) Skylark, The (259)	Bird of the wilderness	V	4, 6, 9 25, 26, 32
47. Holland, Josiah Gilbert (American 1819-1881)			
(a) Christmas Carol, A (226)	There's a song in the air	VI	24, 26, 30
(b) Gradatim (Way to Heaven) (56)	Heaven is not reached at a single bound	VII	1, 4, 6 25, 26
48. Holmes, Oliver Wendell (American 1809-1894)			
(a) Chambered Nautilus, The (26)	This is the ship of pearl which poets feign	VIII	1, 6, 24 25, 26
(b) Last Leaf, The (109)	I saw him once before	VIII	24, 25
(c) Old Ironsides (33)	Ay, tear her tattered ensign down	VI	6, 24, 25 26, 30
(d) Union and Liberty (160)	Flags of the heroes who left us their glory,	VII	14
49. Hood, Thomas (English 1799-1845)			
(a) I Remember, I Remember (227)	I remember, I remember, the house where I was born	IV	4, 6, 25 32

TABLE XV—(*Continued*)

Author and Title	First Line	Grade	Reference
50. Hopkinson, Joseph (American 1770-1842)			
(a) Hail Columbia (152)	Hail Columbia, happy land!	IV	24, 25
51. Houghton, (Lord Milnes, R. W.) (English 1809-1885)			
(a) Good-Night and Good-Morning (81)	A fair little girl sat under a tree	II	9, 32
(b) Lady Moon (82)	Lady moon, lady moon, where are you roving!	II	9, 32
52. Howe, Julia Ward (Mrs.) (American 1819-1910)			
(a) Battle Hymn of the Republic (34)	Mine eyes have seen the glory of the coming of the Lord	VIII	26
53. Howitt, Mary (English 1799-1888)			
(a) Fairies of the Caldon Low, The (292)	And where have you been, my Mary,	III	9, 25, 26 32
(b) Voice of Spring, The (The Coming of Spring) (200)	I am coming, little maiden,	III	9
54. Howlister, Mary (American			
(a) Our Flag (260)	There are many flags in many lands	II	13
55. Hugo, Victor (French 1802-1885)			
(a) Good Night (161)	Good-Night! Good-Night! Far flies the night	I	3
56. Hunt, Leigh (James Henry) (English 1784-1859)			
(a) Abou Ben Adhem (11)	Abou Ben Adhem (may his tribe increase!)	VI	1, 4, 6 9, 25, 26 30
57. Ingelow, Jean (English 1820-1897)			
(a) Seven Times One (18)	There's no dew left on the daisies and clover	II	9, 25, 32
58. Jackson, Helen Hunt (Mrs.) (American 1831-1885)			
(a) Down to Sleep (75)	November woods are bare and still,	V	30
(b) October's Bright Blue Weather (12)	O suns and skies and clouds of June,	IV	26, 30
(c) September (6)	The goldenrod is yellow	III	32
59. Jelliffe, Helena L. (Mrs.) (American 1867-1916)			
(a) Clovers, The (293)	The clovers have no time to play	II	7
60. Jewett, Sarah Orne (American 1849-1909)			
(a) Discontent (182)	Down in the field one day in June	II	32

TABLE XV—(*Continued*)

Author and Title	First Line	Grade	Reference
61. Jones, Wm. (English 1746-1794)			
(a) What Constitutes a State? (261)	What constitutes a state?	VIII	4, 25, 26
62. Keats, John (English 1795-1821)			
(a) On the Grasshopper and Cricket (294)	The poetry of earth is never dead:	V	25, 30
63. Keble, John (English 1792-1866)			
(a) All Things Beautiful (295)	All things bright and beautiful,	I	12
64. Key, Francis Scott (American 1780-1843)			
(a) Star Spangled Banner, The (8)	O say, can you see, by the dawn's early light	V	6, 24, 25 26
65. Kingsley, Chas. (English 1819-1875)			
(a) Farewell, A (162)	My fairest child, I have no song to give you;	IV	4, 6, 9 25, 26, 30
(b) Lost Doll, The (Water Babies, The) (83)	I once had a sweet little doll, dears,	II	9
(c) Three Fishers, The (262)	Three fishers went sailing away to the West	VI	4, 25
66. Kipling, Rudyard (English 1865-)			
(a) If (263)	If you can keep your head when all about you	VIII	1
(b) L'Envoi (228)	When earth's last picture is painted, and and tubes are twisted and dried	VIII	1, 25
(c) Recessional (48)	God of our fathers, known of old	VIII	1, 6, 20 25, 26, 30
(d) White Seal, The (Seal's Lullaby, The) (183)	Oh, hush thee my baby the night is behind us	II	32
67. Krout, Mary Hannah (American 1857-)			
(a) Little Brown Hands (163)	They drive home the cows from the pasture	III	12
68. Lang, Andrew (Scotch 1844-1912)			
(a) Scythe Song (264)	Mowers, weary and brown, and blithe,	VI	25, 30
69. Lanier, Sidney (American 1842-1881)			
(a) Song of the Chattahoochee, The (229)	Out of the hills of Habersham	VIII	1, 24, 25

TABLE XV—(*Continued*)

Author and Title	First Line	Grade	Reference
70. Larcom, Lucy (American 1826-1893)		II	25, 26, 31
(a) Brown Thrush, The (35)	There's a merry brown thrush sitting up in the tree	IV	16
(b) Calling the Violet (296)	Dear little Violet,	II	16
(c) If I Were a Sunbeam (201)	If I were a sunbeam,	III	32
(d) Rivulet, The (297)	Run, little rivulet, run!		
71. Lear, Edward (English 1812-1888)			
(a) Nonsense Alphabet (265)	A was an ant who seldow stood still	I	19
(b) Owl and the Pussy Cat, The (115)	The owl and the pussy cat went to sea	II	1, 6, 25 32
72. Longfellow, Henry Wadsworth (American 1807-1882)			
(a) Arrow and the Song, The (13)	I shot an arrow into the air,	IV	6, 18, 24 25, 26
(b) Bell of Atri, The (202)	At Atri in Abruzzo, a small town	IV	18
(c) Builders, The (57)	All are architects of Fate,	VI	1, 18
(d) Building of the Ship, The (153)	Then the master with a gesture of command	VII	1, 18
(e) Children's Hour, The (2)	Between the dark and the daylight	III	1, 4, 18 24, 25, 26
(f) Courtship of Miles Standish, The (266)	In the Old Colony days, in Plymouth the land of the Pilgrims	VII	18
(g) Day is Done, The (44)	The day is done, and the darkness	V	1, 18, 25
(h) Daybreak (103)	A wind came up out of the sea	IV	18, 25
(i) Evangeline (203)	This is the forest primeval	VII	18, 24
(j) Excelsior (140)	The shades of night were falling fast	V	18, 25, 26
(k) Hiawatha (68)	Should you ask me whence these stories?	II	18
(l) Hiawatha's Childhood (36)	By the shores of Gitche Gumee	II	6, 9, 18 25
(m) Hiawatha's Friends (298)	Two good friends of Hiawatha	IV	18
(n) Hiawatha's Sailing (104)	Give me your bark, O Birch Tree!	III	18
(o) King Robert of Sicily (299)	Robert of Sicily, brother of Pope Urbane	VII	18
(p) Ladder of St. Augustine, The (267)	Saint Augustine! well hast thou said	VII	18
(q) Old Clock on the Stairs, The (204)	Somewhat back from the village street,	V	18, 25
(r) Paul Revere's Ride (49)	Listen my children, and you shall hear	V	1, 18, 25 26

TABLE XV—(*Continued*)

Author and Title	First Line	Grade	Reference
(s) Psalm of Life, A (63)	Tell me not in mournful numbers	VI	1, 6, 18 24, 25, 26
(t) Rain in Summer (131)	How beautiful is the rain	IV	18, 30
(u) Rainy Day, The (164)	The day is cold, and dark, and dreary,	VII	18, 25
(v) Sandalphon (300)	Have you read in the Talmud of old?	VI	18
(w) Ship of State, The (Building of the Ship, The) (105)	Thou, too, sail on, O ship of State	V	18
(x) Village Blacksmith, The (4)	Under a spreading Chestnut tree	IV	6, 9, 18 24, 25, 26
(y) Wreck of the Hesperus, The (205)	It was the schooner Hesperus	IV	6, 9, 18 25, 26, 30
73. Lowell, James Russell (American 1819-1891)			
(a) Aladdin (230)	When I was a beggarly boy,	V	25
(b) Day in June, A (132)	O, what is so rare as a day in June	VIII	1, 4, 6 24, 25, 26
(c) Fatherland, The (133)	Where is the true man's fatherland	VII	25, 26, 30
(d) Finding of the Lyre (231)	There lay upon the ocean's shore	VII	6
(e) First Snowfall, The (15)	The snow had begun in the gloaming	IV	24, 25
(f) Fountain, The (141)	Into the sunshine	IV	12
(g) Heritage, The (206)	The rich man's son inherits land	V	25, 26
(h) Ode Recited at the Harvard Commemoration (301)	Weak winged is song,	VIII	24, 25
(i) Stanzas on Freedom (268)	They are slaves who fear to speak	VIII	24
(j) To the Dandelion (165)	Dear common flower that grows beside the way,	VI	25, 26, 30
(k) Vision of Sir Launfal, The (69)	Over his keys the musing organist	VIII	12
(l) Yussouf (207)	A stranger came one night to Yussouf's tent	VII	12
74. Macaulay, Thomas Babington (English 1800-1859)			
(a) Horatius (Lays of Ancient Rome (184)	Lars Porsena of Clusium	VII	1, 6, 25 26
75. MacDonald, George (Scotch 1824-1905)			
(a) Baby, The (At the Back of the North-wind) (94)	Where did you come from baby dear?	I	1, 24, 26
(b) Little White Lily, The (269)	Little white lily	II	6, 25, 26 32
(c) Wind and the Moon, The (142)	Said the wind to the moon, "I will blow you out."	IV	6, 25, 26

TABLE XV—(*Continued*)

Author and Title	First Line	Grade	Reference
76. Mackay, Chas, (Scotch 1814-1889)			
(a) Miller of the Dee, The (270)	There dwelt a miller hale and bold	III	24, 25
(b) Tubal Cain (232)	Old Tubal Cain was a man of might	VI	12
77. Miller, Emily (Mrs.) (American 1833-1913)			
(a) Bluebird, The (30)	I know the song that the bluebird is singing	II	32
78. Miller, Joaquin (Miller, Cincinatus Hiner) (American 1841-1913)			
(a) Columbus (45)	Behind him lay the gray Azores	VII	1, 6, 24 25, 26, 30
79. Milton, John (English 1608-1674)			
(a) On His Blindness (302)	When I consider how my light is spent	VIII	1, 4, 6 25, 26
(b) Song on a May Morning (303)	Now the bright morning star, day's harbinger	V	25, 26, 30
80. Moore, Clement Clarke (American 1779-1863)			
(a) Visit from St. Nicholas, A (40)	'Twas the night before Christmas, when all through the house,	III	24, 26
81. Moore, Thomas (Irish 1779-1852)			
(a) Minstrel Boy, The (304)	The Minstrel-boy to the war has gone,	VI	25, 30
82. Morris, George Pope (American 1802-1864)			
(a) Woodman, Spare That Tree (106)	Woodman, spare that tree	V	4, 6, 24 25
83. Nesbitt, Wilbur Dick (American 1871-)			
(a) Your Flag and My Flag (271)	Your flag and my flag	III	12
84. Newman, John Henry (Cardinal Newman) (English 1801-1890)			
(a) Lead Kindly Light (Pillar of the Cloud, The) (166)	Lead kindly light, amid the encircling gloom	VIII	4, 6, 25
85. Payne, John Howard (American 1791-1852)			
(a) Home, Sweet Home (76)	'Mid pleasures and palaces though we may roam,	IV	6, 25
86. Perry, Nora (American 1832-1896)			
(a) Coming of Spring, The (167)	There's something in the air	VI	26

TABLE XV—(*Continued*)

Author and Title	First Line	Grade	Reference
87. Pierpont, John (American 1785-1866)			
(a) Warren's Address (84)	Stand! the ground's your own, my braves!	VIII	4, 6, 24 25, 26, 30
88. Poulsson, Emilie (American 1853-)			
(a) First Christmas, The (305)	Once a little baby lay	I	13
(b) While Stars of Christmas Shine (306) .	While stars of Christmas shine	I	13
89. Proctor, Adelaide (English 1825-1864)			
(a) One by One (272)	One by one the sands are flowing,	VI	25
90. Procter, Bryan Waller (Barry Cornwall) (English 1787-1874)			
(a) Sea, The (116)	The sea! the sea! the open sea	IV	24, 25, 26 30
91. Rands, William Bright (English 1823-1882)			
(a) Great, Wide, Beautiful, Wonderful, World (Child's World, The; Wonderful World, The) (23)	Great, wide, beautiful, wonderful world	III	24, 25, 26 32
92. Read, Thomas Buchanan (American 1822-1872)			
(a) Sheridan's Ride (154)	Up from the south, at break of day,	VI	1, 6, 25 26
93. Riley, James Whitcomb (American 1853-1916)			
(a) Brook Song, A (208)	Little brook! Little brook!	IV	21
(b) Knee Deep in June (185)	Tell you what I like the best	V	1
(c) Life Lesson, A (233)	There! little girl, don't cry!	VII	24, 25
(d) Little Orphant Annie (155)	Little orphant Annie's come to our house to stay,	III	26
(e) Old Glory (Name of Old Glory, The (95)	Old Glory! say, who by the ships and the crew	VI	30
(f) Song, A (168)	There is ever a song somewhere my dear	V	21
(g) Sudden Shower, A (143)	Barefooted boys scud up the street	IV	21
94. Rossetti, Christina Georgina (English 1830-1894)			
(a) Boats Sail on the Rivers (96)	Boats sail on the rivers	I	12
(b) Milking Time (273)	When the cows come home, the milk is coming	II	32

TABLE XV—(*Continued*)

Author and Title	First Line	Grade	Reference
(c) O, Lady Moon (307)	O Lady Moon, your horns point toward the east:	I	32
(d) Swallow, The (308)	Fly away, fly away, over the sea,	III	19
(e) What Does the Bee do? (309)	What does the bee do?	I	12
(f) Wind, The (58)	Who has seen the wind?	I	25, 26, 31
95. Sangster, Margaret Elizabeth (American 1838-1912)			
(a) We Thank Thee (310)	For peace and for plenty, for freedom, for rest,	II	2
96. Saxe, John Godfrey (American 1816-1887)			
(a) Blind Men and the Elephant, The (311)	It was six men of Indostan	V	25
97. Scott, Sir Walter (Scotch 1771-1832)			
(a) Breathes There a Man (Love of Country; My Native Land; Patriotism) (7)	Breathes there a man with soul so dead	VIII	1, 4, 9 26, 30
(b) Lady of the Lake (312)	Harp of the north that mouldering long has hung	VIII	22
(c) Lochinvar (209)	Oh, young Lochinvar's come out of the west	VI	6, 25, 26 30
(d) Lullaby of an Infant Chief (234)	O, hush thee, my babie, thy sire was a knight,	II	25, 32
98. Shakespeare, Wm. (English 1564-1616)			
(a) Ariel's Song (Tempest, The) (313)	I. Come unto these yellow sands	II	30
	II. Foot it featly here and there;		30
	III. Where the bee sucks there suck I;		9, 30
	IV. Full fathom five thy father lies;		30
(b) Good Name (Reputation) (Othello) (210)	Good name in man and woman, dear my lord,	VII	
(c) Hark, Hark! the Lark (Cymbeline) (186)	Hark! Hark! the lark at Heaven's gate sings	VI	4, 9, 25 30
(d) Mark Anthony's Address (Julius Caesar) (235)	Friends, Romans, Countrymen!	VIII	6
(e) Mercy Speech (Merchant of Venice) (50)	The quality of mercy is not strained	VIII	1, 6
(f) Orpheus With His Lute (King Henry the Eighth) (236)	Orpheus with his lute made trees,	VI	9, 30
(g) Polonius to Laertes (Hamlet) (144)	And these few precepts in thy memory	VIII	1, 6, 30
(h) Under the Greenwood Tree (As You Like It) (117)	Under the greenwood tree	V	9, 25, 26 30

TABLE XV—(*Continued*)

Author and Title	First Line	Grade	Reference
99. Shaw, D. T. (American)			
(a) Columbia, the Gem of the Ocean (314)	O, Columbia the gem of the ocean	III	12
100. Shelley, Percy, Bysshe (English 1792-1822)			
(a) Cloud, The (118)	I bring fresh showers for the thirsty flowers	VIII	1, 25, 30
(b) To a Skylark (169)	Hail to thee, blithe spirit	VIII	1, 4, 6 25, 26, 30
101. Sherman, Frank Dempster (American 1860-1916)			
(a) Clouds (187)	The sky is full of clouds today	III	23, 32
(b) Daisies (37)	At evening when I go to bed	II	12, 23
(c) Dewdrop, A (119)	Little drop of dew	I	23, 31
(d) Four Winds, The (110)	In winter, when the wind I hear	II	23
(e) Goldenrod (156)	Spring is the morning of the year	I	23
(f) Hide and Seek (238)	Now hide the flowers beneath the snow,	II	23
(g) Leaves at Play (170)	Scamper, little leaves, about	I	23
(h) May (274)	May shall make the world anew,	VI	23, 32
(i) Real Santa Claus, A (237)	Santa Claus I hang for you	III	23
(j) Snowbird, The (239)	When all the ground with snow is white,	I	23
(k) Snowflakes (211)	Out of the sky they come	II	23
(l) Song for Winter (315)	Now winter fills the world with snow	I	23
(m) Wizard Frost (212)	Wonderous things have come to pass	III	23
102. Sill, Edward Roland (American 1841-1887)			
(a) Opportunity (111)	This I beheld, or dreamed it in a dream:—	VIII	1, 25, 30
103. Smith, Samuel Francis (American 1808-1895)			
(a) America (1)	My country 'tis of thee	II	6, 24, 25 26
104. Southey, Caroline (Mrs.) (Bowles, Ann) (English 1786-1854)			
(a) Little Ladybird, The (316)	Lady-bird! lady-bird fly away home	III	31
105. Southey, Robert (English 1774-1843)			
(a) Inchcape Rock, The (213)	No stir in the air, no stir in the sea,	V	6, 9, 25 30

TABLE XV—(*Continued*)

Author and Title	First Line	Grade	Reference
106. Stedman, Edmund Clarence (American 1833-1908)			
(a) What the Winds Bring (188)	Which is the wind that brings the cold?	III	32
107. Stevenson, Robert Louis (Scotch 1850-1894)			
(a) Autumn Fires (120)	In the other gardens	I	27
(b) Bed in Summer (14)	In the winter I get up at night	I	9, 27
(c) Cow, The (77)	The friendly cow all red and white,	I	9, 31
(d) Farewell to the Farm (318)	The coach is at the door at last;	II	27
(e) Foreign Children (171)	Little Indian, Sioux or Crow	II	27
(f) Foreign Lands (172)	Up into the cherry tree	II	25, 26, 27
(g) Good Play, A (317)	We built a ship upon the stairs	II	27
(h) Happy Thought (275)	The world is so full of a number of things,	I	25, 26, 27
(i) Lamplighter, The (145)	My tea is nearly ready and the sun has left the sky	I	9, 27
(j) Land of Counterpane, The (121)	When I was sick and lay abed	I	9, 26, 27
(k) Land of Story Books, The (38)	At evening when the lamp is lit	II	26, 27, 32
(l) Moon, The (320)	The moon has a face like the clock on the wall	I	27
(m) My Bed is a Boat (134)	My bed is like a little boat;	II	25, 26, 27
(n) My Shadow (5)	I have a little shadow that goes in and out with me,	I	6, 24, 25 26, 27
(o) Rain (64)	The rain is raining all around,	I	27, 31
(p) Singing (319)	Of speckled eggs the birdie sings	I	27
(q) Sun's Travels, The (135)	The sun is not abed, when I	II	27
(r) Swing, The (22)	How do you like to go up in a swing,	I	27
(s) Time to Rise (321)	A birdie with a yellow bill	I	27
(t) Where Go the Boats? (65)	Dark brown is the river	II	27, 32
(u) Whole Duty of Children (276)	A child should always say what's true	I	25, 26, 27
(v) Wind, The (9)	I saw you toss the kites on high	II	26, 27, 30
(w) Windy Nights (70)	Whenever the moon and stars are set	II	27, 32
108. Stoddard, Richard Henry (American 1825-1903)			
(a) Abraham Lincoln (322)	This man whose homely face you look upon,	VII	24, 30

TABLE XV—(*Continued*)

Author and Title	First Line	Grade	Reference
109. Tate, Nahum (English 1652-1715)			
(a) Christmas (While Shepherds Watched) (173)	While shepherds watched their flocks by night	IV	25, 26, 30
110. Taylor, Bayard (American 1825-1878)			
(a) Song of the Camp, The (240)	"Give us a Song!" the soldier cried,	VII	6, 24, 25 26, 30
111. Taylor, Jane (English 1783-1824)			
(a) I Like Little Pussy (146)	I like little pussy	I	25, 26, 31
(b) Thank You, Pretty Cow (214)	Thank you, pretty cow, that made	I	32
(c) Twinkle, Twinkle, Little Star (59)	Twinkle, twinkle, little star;	I	31
(d) Violet, The (147)	Down in a green and shady bed	III	25, 26, 32
112. Tennyson, Lord Alfred (English 1809-1892)			
(a) Bird and the Baby, The (25)	What does little birdie say	I	9, 25, 26
(b) Break, Break, Break (136)	Break, break, break,	VI	4, 25
(c) Brook, The (46)	I come from haunts of coot and hern,	IV	6, 25, 26 30
(d) Bugle Song (41)	The splendor falls on castle walls	VI	12
(e) Charge of the Light Brigade (66)	Half a league, half a league,	VI	1, 6, 9 25, 26, 30
(f) Crossing the Bar (137)	Sunset and evening star,	VII	1, 6, 20 25, 26
(g) Eagle, The (323)	He clasps the crag with crooked hands;	VI	25
(h) Flower in the Crannied Wall (215)	Flower in the crannied wall	VIII	28
(i) In Memoriam (241)	Strong son of God, immortal Love,	VII	25
(j) King Arthur (242)	Leodogran the King of Camdiard	VIII	28
(k) Owl, The (157)	When cats run home and light is come,	III	32
(l) Ring Out Wild Bells (In Memoriam) (174)	Ring out wild bells, to the wild sky	VI	28
(m) Sir Galahad (243)	My good blade carves the casques of men,	VII	6, 25
(n) Sweet and Low (19)	Sweet and low, sweet and low,	III	9, 32
(o) Throstle, The (216)	Summer is coming, summer is coming	III	25, 26
(p) Winter, (Window, The) (244)	The frost is here	IV	28
113. Thaxter, Celia (Mrs.) (Alia Leighton) (American 1835-1894)			
(a) April (324)*		II	

---

*There seems to some doubt whether "April" and "Spring" are different poems or not. Granger mentions both titles. However, we have found the poem whose first line is "The alder by the river" under both titles.

TABLE XV—(*Continued*)

Author and Title	First Line	Grade	Reference
(b) Chanticleer (277)	I wake! I feel the day is near;	II	32
(c) Little Gustava (217)	Little Gustava sits in the sun,	III	25, 26, 32
(d) March (158)	I wonder what spendthrift chose to spill	III	13
(e) Piccola (325)	Poor, sweet Piccola! did you hear	V	13
(f) Sandpiper, The (31)	Across the narrow beach we flit,	IV	6, 24, 25 26, 30
(g) Spring (97)*	The alder by the river	II	13, 32
(h) Wild Geese (189)	The wind blows, the sun shines, the birds sing loud	II	12
114. Thomas, Edith Matilda (American 1854-)			
(a) Talking in Their Sleep (190)	You think I am dead,	III	12
115. Vandegrift, Margaret (Janvier Margaret) (American 1845-1913)			
(a) Sand Man, The (191)	The rosy clouds float overhead,	III	25, 26, 3
116. VanDyke, Henry (American 1852-)			
(a) Song Sparrow, The (326)	There is a bird I know so well,	III	12
117. Wadsworth, Olive A. (Mrs.) (Dana, Katharine Floyd) (American 1835-1886)			
(a) Over in the Meadow (112)	Over in the meadow,	I	12
118. Watts, Isaac (English 1674-1748)			
(a) Busy Bee, The (245)	How doth the little busy bee	I	9
119. Whitman, Walt (American 1819-1892)			
(a) O Captain! My Captain! (60)	O Captain! My Captain! our fearful trip is done,	VIII	1, 6, 9 24, 25, 26 30
120. Whittier, John Greenleaf (American 1807-1892)			
(a) Barbara Fritchie (138)	Up from the meadows rich with corn,	V	6, 9, 25 26, 29, 30
(b) Barefoot Boy, The (20)	Blessings on three, little man,	V	1, 24, 25 26, 29, 30
(c) Corn Song, The (42)	Heap high the farmer's wintry hoard!	V	29, 30
(d) Huskers, The (246)	It was late in mild October, and the long autumnal rain	VII	29
(e) In School Days (88)	Still sits the schoolhouse by the road,	V	24, 29
(f) Snowbound: A Winter Idyll (85)	Unwarmed by any sunset light	VIII	24, 29, 30

*See note regarding Thaxter's "April."

TABLE XV—(*Continued*)

(g) Three Bells, The (327)	Beneath the low hung night cloud	IV	29
121. Wilder, J. N. (American)			
(a) Stand by the Flag (278)	Stand by the Flag! Its stars, like meteors gleaming,	VII	1
122. Wolfe, Chas. (English 1791-1823)			
(a) Burial of Sir John Moore, The (192)	Not a drum was heard, not a funeral note	VI	4, 6, 9 25, 26, 30
123. Wordsworth, Wm. (English 1770-1850)			
(a) Fidelity (328)	A barking sound the shepherd hears,	VI	6
(b) I Wandered Lonely as a Cloud (Daffodils, The) (21)	I wandered lonely as a cloud	VI	1, 4, 6 25, 26, 32
(c) Kitten and the Falling Leaves, The (248)	That way look, my infant lo!	III	26, 32
(d) Lucy Gray (Solitude)... (247)	Oft I had heard of Lucy Gray;	IV	9, 25, 26 32
(e) My Heart Leaps up When I Behold (Rainbow, The) (218)	My heart leaps up when I behold	V	4, 6, 25 26
(f) Pet Lamb, The (329)	The dew was falling fast, the stars began to blink;	IV	9
(g) To a Butterfly (148)	I've watched you now a full half hour,	III	25
(h) To a Skylark (193)	Up with me, up with me into the clouds	VI	25, 26
(i) We Are Seven (279)	A simple child, (I met a little cottage girl)	IV	26
(j) Written in March (175)	The cock is crowing	IV	4, 25, 26

We have tried to make the number of titles in this bibliography as small as possible, omitting collections which contained but one or two of the poems in our list. The result is that a few poems are not in any collection listed in the bibliography. Some of these are to be found in the collected works of poets—e.g. Tennyson's "The Bugle Song," Shakespeare's "Good Name" (Othello), etc. Others are printed in readers—such as Wadsworth's "Over in the Meadow." There are still a few poems, however, which are not to be found in any of these sources. In such cases we have referred the reader to Granger's "Index to Poetry and Recitation" (Code No. 12). This book does not contain the text of these poems but it gives references to sources in which the text appears.

IX—AUTHORS MENTIONED IN THE SELECTED LIST

Among the questions brought up by this study are the following: Who are the authors whose poems are mentioned in the Selected List? Who are the most popular authors? To what grades are their poems best adapted? Tables XVI and XVII attempt to answer these questions.

One hundred twenty-two poets contributed 326 of the 329 poems in the Selected List. The remaining three poems are anonymous. The list of authors contains the names of most of the major English and American poets, as well as the names of many minor poets—particularly of the minor American poets. One of the interesting possibilities of this study would be an attempt to find out how nearly the material included in the Selected List approximates the type of literature with which our children should become familiar. While this is a problem for literary criticism and is quite outside the scope of the present report, one may at least point out two dominant tendencies as exhibited by the Selected List. The first is the marked tendency to favor things American. Not only is this apparent in the patriotic selections so often referred to, but it is also clear from the nationality of the authors. Of the 326 poems whose authors are known 183 are by American authors. This is a much larger number than the body of high-grade American and English verse would lead us to expect. It seems clear that the inclusion of an American poem offers less conclusive evidence of its merit than does the inclusion of an English poem.

TABLE XVI—NUMBER OF POEMS BY AUTHORS AND BY SUGGESTED GRADES. SELECTED LIST

Author	Primary Grades (I, II)	Intermediate Grades (III, IV, V)	Grammar Grades (VI, VII, VIII)	Total
Addison, Joseph			1	1
Aldrich, T. B.		1	1	2
Alexander, Mrs. C. F.	1		1	2
Allingham, Wm.	2	1		3
Allison, Joy	1			1
Anonymous*	2	1		3
Bangs, J. K.	1			1
Bates, Clara D.	1			1
Bennett, H. H.		1		1
Bjornson, Bjornsterne		1		1
Blake, Wm.	1		1	2
Brooks, Phillips		1	1	2

*A Laughing Chorus; The Secret; Sleep, Baby Sleep—From the German.

TABLE XVI—(*Continued*)

Author	Primary Grades (I, II)	Intermediate Grades (III, IV, V)	Grammar Grades (VI, VII, VIII)	Total
Brown, Kate L.	1			1
Browning, Elizabeth B.		1		1
Browning, Robert		1	4	5
Bryant, W. C.		3	7	10
Bunner, H. C.	1			1
Burns, Robert			3	3
Byron, Lord			2	2
Campbell, Thomas			2	2
Carlyle, Thomas		1		1
Cary, Alice		1	2	3
Cary, Phoebe	3	2		5
Child, Lydia M.	2			2
Coleridge, S. T.	1	1		2
Cone, Helen G.		1		1
Coolidge, Susan	1			1
Cooper, George	3	1		4
Cowper, Wm.		1		1
Deland, Mary		1		1
Dodge, Mary	1			1
Drake, J. R.			1	1
Edwards, Matilda B.		1		1
Emerson, R. W.	1	2	4	7
Field, Eugene	7	3		10
Fields, J. T.		1		1
Finch, F. M.			1	1
Follen, Eliza	1			1
Garabrant, Nellie M.	1			1
Gould, Hannah F.		1		1
Gray, Thomas			1	1
Hale, Sarah J.	1			1
Hemans, Felicia		2		2
Herford, Oliver	1			1
Higginson, Ella		1		1
Hogg, James		2		2
Holland, J. G.			2	2
Holmes, O. W.			4	4
Hood, Thomas		1		1
Hopkinson, Joseph		1		1
Houghton, Lord	2			2
Howe, Julia W.			1	1
Howitt, Mary		2		2
Howlister, Mary	1			1
Hugo, Victor	1			1
Hunt. Leigh	1			1
Ingelow, Jean	1			1
Jackson, Helen H.		3		3
Jelliffe, Helena L.	1			1
Jewett, Sarah O.	1			1
Jones, Wm.			1	1
Keats, John		1		1
Keble, John	1			1
Key, F. S.		1		1
Kingsley, Charles	1	1	1	3
Kipling, Rudyard	1		3	4
Krout, Mary		1		1
Lang, Andrew			1	1

TABLE XVI—(*Continued*)

Author	Primary Grades (I, II)	Intermediate Grades (III, IV, V)	Grammar Grades (VI, VII, VIII)	Total
Lanier, Sidney			1	1
Larcom, Lucy	2	2		4
Lear, Edward	2			2
Longfellow, H. W.	2	14	9	25
Lowell, J. R.		4	8	12
Macaulay, T. B.			1	1
MacDonald, George	2	1		3
Mackay, Charles		1	1	2
Miller, Emily	1			1
Miller, Joaquin			1	1
Milton, John		1	1	2
Moore, C. C.		1		1
Moore, Thomas			1	1
Morris, G. P.		1		1
Nesbitt, W. D.		1		1
Newman, J. H.			1	1
Payne, J. H.		1		1
Perry, Nora			1	1
Pierpont, John			1	1
Poulsson, Emilie	2		1	1
Procter, Adelaide			1	1
Procter, B. W.		1		1
Rands, W. B.		1		1
Read, T. B.			1	1
Riley, J. W.		5	2	7
Rossetti, Christina	5	1		6
Sangster, Margaret	1			1
Saxe, J. G.		1		1
Scott, Sir Walter	1		3	4
Shakespeare, Wm.	1	1	6	8
Shaw, D. T.		1		1
Shelley, P. B.			2	2
Sherman, F. D.	9	3	1	13
Sill, E. R.			1	1
Smith, S. F.	1			1
Southey, Caroline		1		1
Southey, Robert		1		1
Stedman, Edmund		1		1
Stevenson, R. L.	23			23
Stoddard, R. H.			1	1
Tate, Nahum		1		1
Taylor, Bayard			1	1
Taylor, Jane	3	1		4
Tennyson, Alfred	1	5	10	16
Thaxter, Celia	4	4		8
Thomas, Edith		1		1
Vandegrift, Margaret		1		1
Van Dyke, Henry		1		1
Wadsworth, Olive A.	1			1
Watts, Isaac	1			1
Whitman, Walt			1	1
Whittier, J. G.		5	2	7
Wilder, J. N.			1	1
Wolfe, Charles			1	1
Wordsworth, Wm.		7	3	10

Who are Excluded—The second tendency is evident in the practical exclusion of all foreign writers. Of the 329 poems only two, "Sleep Baby Sleep" (anonymous) and Hugo's "Good Night," are translations from another language. It is a question whether some of the admirable French and German poems—some of which have been translated by poets shown in this study to be especially popular—should not be included. It may be added that one or two of the national hymns of other countries just missed inclusion in the Selected List. The Marseillaise was mentioned *four* times.

An examination of the authors and poems listed in our tables reveals some surprising omissions. For example, even among American poets we do not find Poe. We should have expected to find at least two or three of his poems—the ones especially characterized by richness of imagery and rhythm. Among English poets no mention is made of Mathew Arnold nor of Dryden. Keats is represented only by "On the Grasshopper and the Cricket"—a questionable selection. Four lines from the "Ancient Mariner" and the "Answer to a Child's Question" are all that we find by Coleridge. As has been pointed out before, there is a total absence of modern verse. A careful selection of recent poems should be made, and a portion of it should be taught, for some of the poetry of the day is undoubtedly good.

Who are Included—One hundred twenty-two poets, however—a sufficient number, one might say—*are* represented on our Selected List of poems. The following table shows the number of poets who contributed the indicated number of poems as shown by Table XVI.

Number of Poems	Number of Poets
1	76
2	18
3	6
4	6
5	2
6	1
7	3
8	2
9	0
10	3
11	0
12	1
13	1
16	1
23	1
25	1
Total	122

Thus, 76 or more than half of the 122 poets contributed but one poem to our list. Only eight poets, Longfellow, Stevenson, Tennyson, Sherman, Lowell, Field, Bryant, and Wordsworth had ten or more poems in the Selected List. The poets who contributed six or more poems are as follows:

Longfellow	25
Stevenson	23
Tennyson	16
Sherman	13
Lowell	12
Field	10
Bryant	10
Wordsworth	10
Shakespeare	8
Thaxter	8
Emerson	7
Riley	7
Whittier	7
Rossetti (Christina)	6

Grades for Which Each Author's Poems are Appropriate—Table XVI shows the grading of the poems contributed by each author. This is shown not by yearly grades but by groups of yearly grades. Grades I and II form the first group; Grades III -V the second group; and Grades VI -VIII the third. This division of the elementary school was used because it was thought that the interest and abilities of children in these groups were rather sharply defined. These divisions are commonly referred to as Primary, Intermediate and Grammar grades. The idea was to indicate the authors whose poems as shown in our lists were most suitable to the above divisions of the elementary school. For example all of Stevenson's poems are listed for the Primary grades. Eugene Field and Frank Dempster Sherman also wrote for the most part for the Primary children. Longfellow is clearly the poet of the intermediate grades, although he contributed a number of poems for the grammar grades. Most of Tennyson's and Lowell's poetry was regarded as suitable for the grammar grades although each contributed a few poems to lower grades.

Popularity of Authors—In order to determine the relation between the number of poems which an author furnished to the Selected List, and the popularity of these poems Table XVII was devised. It indicates the frequency of mention (all poems considered) which each author received in the Primary, Intermediate and Grammar grades, as well as the total frequency for all grades combined. The table reads as follows (second line): Aldrich was mentioned by the listing of a poem or poems four times in Grades I, and II, seventeen times in Grades III, IV, and V, and eleven times in Grades VI, VII, and VIII—a total of thirty-two listings. This table may be used in the same way as Table XVI to find out the poets best suited to Primary, Intermediate, or Grammar grades.

TABLE XVII—FREQUENCY OF MENTION OF EACH AUTHOR REPRESENTED IN THE SELECTED LIST

Author	Primary Grades (I, II)	Intermediate Grades (III, IV, V)	Grammar Grades (VI, VII, VIII)	Total
Addison, Joseph		1	12	13
Aldrich, T. B.	4	17	11	32
Alexander, Mrs. C.	18	3	4	25
Allingham, Wm.	22	30	1	53
Allison, Joy	12	4		16
Anonymous*	24	4		28
Bangs, J. K.	6			6
Bates, Clara D.	6	1		7
Bennett, H. H.	2	19	12	33
Bjornson, Bjornsterne	8	12		20
Blake, Wm.	8	5	3	16
Brooks, Phillips	3	23	7	33
Brown, Kate L.	21	1		22
Browning, Elizabeth B.	1	8		9
Browning, Robert	9	20	37	66
Bryant, W. C.	1	65	80	146
Bunner, H. C.	13	10		23
Burns, Robert		4	22	26
Byron, Lord			11	11
Campbell, Thomas		1	13	14
Carlyle, Thomas	9	13	1	23
Cary, Alice		22	25	47
Cary, Phoebe	22	27	2	51
Child, Lydia M.	31	13	1	45
Coleridge, S. T.	6	17	2	25
Cone, Helen		7	2	9
Coolidge, Susan	14	9		23
Cooper, George	30	9		39
Cowper, Wm.		4	1	5
Deland, Mary		5		5
Dodge, Mary	5			5
Drake, J. R.	2	7	19	28
Edwards, Matilda B.	4	4		8
Emerson, R. W.	11	33	46	90
Field, Eugene	130	87	1	218
Fields, J. T.	1	9	1	11
Finch, F. M.		2	15	17
Follen, Eliza	13			13
Garabrant, Nellie	5	3	1	9
Gould, Hannah F.	6	13		19
Gray, Thomas			9	9
Hale, Sarah J.	6			6
Hemans, Felicia		25	14	39

*A Laughing Chorus; The Secret; Sleep, Baby Sleep—From the German.

TABLE XVII—(*Continued*)

Author	Primary Grades (I, II)	Intermediate Grades (III, IV, V)	Grammar Grades (VI, VII, VIII)	Total
Herford, Oliver	7	1		8
Higginson, Ella	2	6		8
Hogg, James	2	24	3	29
Holland, J. G.	1	8	21	30
Holmes, O. W.		8	74	82
Hood, Thomas		5	2	7
Hopkinson, Joseph	2	5	4	11
Houghton, Lord	29	9		38
Howe, Julia W.	2	7	18	27
Howitt, Mary	1	12		13
Howlister, Mary	3	3		6
Hugo, Victor	9	1		10
Hunt, Leigh		13	22	35
Ingelow, Jean	22	11		33
Jackson, Helen H.	17	73	6	96
Jelliffe, Helena L.	5			5
Jewett, Sarah O.	6	3		9
Jones, Wm.			6	6
Keats, John		3	2	5
Keble, John	5			5
Key, F. S.	4	20	14	38
Kingsley, Chas.	16	14	5	35
Kipling, Rudyard	3	4	39	46
Krout, Mary		10		10
Lang, Andrew			6	6
Lanier, Sidney		1	6	7
Larcom, Lucy	24	21		45
Lear, Edward	12	8		20
Longfellow, H. W.	51	232	133	416
Lowell, J. R.	1	58	85	144
Macaulay, T. B.		1	8	9
MacDonald, George	24	11		35
Mackay, Charles		7	6	13
Miller, Emily	15	13		28
Miller, Joaquin		3	22	25
Milton, John		3	7	10
Moore, C. C.	9	17		26
Moore, Thomas		1	4	5
Morris, G. P.		14	2	16
Nesbitt, W. D.	1	4	1	6
Newman, J. H.		3	7	10
Payne, J. H.	2	12	6	20
Perry, Nora	2	4	4	10
Pierpont, John		3	16	19
Poulsson, Emilie	8	2		10
Procter, Adelaide		3	3	6
Procter, B. W.		9	5	14
Rands, W. B.	10	22		32
Read, T. B.		1	10	11
Riley, J. W.		49	25	74
Rossetti, Christina G.	56	5		61
Sangster, Margaret	4	1		5
Saxe, J. G.		3	2	5
Scott, Sir Walter	3	13	44	60
Shakespeare, W.	3	21	62	86
Shaw, D. T.	1	4		5

TABLE XVII—(*Continued*)

Author	Primary Grades (I, II)	Intermediate Grades (III, IV, V)	Grammar Grades (VI, VII, VIII)	Total
Shelley, P. B.	1	4	19	24
Sherman, F. D.	96	35	3	134
Sill, E. R.			15	15
Smith, S. F.	26	25	8	59
Southey, Caroline	2	3		5
Southey, Robert		5	3	8
Stedman, Edmund	3	6		9
Stevenson, R. L.	323	58	1	382
Stoddard, R. H.		1	4	5
Tate, Nahum		7	3	10
Taylor, Bayard		1	6	7
Taylor, Jane	44	11		55
Tennyson, A.	46	78	109	233
Thaxter, Celia	33	53	3	89
Thomas, Edith M.	2	7		9
Vandegrift, Margaret	1	8		9
Van Dyke, Henry	1	3	1	5
Wadsworth, Olive A.	15			15
Watts, Isaac	7			7
Whitman, Walt			23	23
Whittier, J. G.	3	81	37	121
Wilder, J. N.		2	4	6
Wolfe, Charles		1	8	9
Wordsworth, W.	7	61	34	102

The fourteen authors who received the highest frequency of mention are the following:

Author	Frequency of Mention
Longfellow	416
Stevenson	382
Tennyson	233
Field	218
Bryant	146
Lowell	144
Sherman	124
Whittier	121
Wordsworth	102
Jackson	96
Emerson	90
Thaxter	89
Shakespeare	86
Holmes	82

Twelve of these fourteen authors are also on the list of the fourteen who contributed the greatest number of poems to the Selected List. Longfellow, Stevenson, and Tennyson are the first three poets on each list in the order given. The first seven poets on both lists are likewise the same, but there is some variation in their order.

The two poets who were on the list of the fourteen receiving the highest frequency of mention, but who were not on the list of the fourteen who contributed the most poems were Holmes and Jackson. Only four of Holmes' poems are included in the Selected List; but each of them received a high frequency of mention. Likewise the three poems contributed by Helen Hunt Jackson were very popular. Thus the frequency of mention which she received was greater than that of several authors who contributed a larger number of less popular poems. The authors who contributed a large enough number of poems to place them among the fourteen having the largest number of poems mentioned, but whose poems were not mentioned often enough to place them among the fourteen most frequently mentioned, were Rossetti and Riley. Rossetti contributed seven poems, only one of which received a mention as high as seventeen, and three of which received a mention of five (the lowest number of "mentions" which would secure the listing of a poem). Riley, who contributed seven poems to the Selected List, ranked fifteenth in "Frequency of Mention" and consequently fell just outside the group of fourteen most frequently mentioned authors.

Patriotic poems and well-known selections by authors who contributed but one or two poems inversely affected the relationship between the two tables. For instance, "America" with a total frequency of 59 gives S. F. Smith a comparatively high rank in the list of "Total Frequency of Mention" while he is at the foot of the list of "Number of Poems Contributed." The same condition holds true for Key's "Star Spangled Banner" with a frequency of 38. Ingelow also contributed one popular poem "Seven Times One" with a frequency of 33; and "The American Flag", the only poem by Drake in our list, has a frequency of mention of 28.

These cases, however, were so infrequent that the correlation was effected but slightly. The closeness of correspondence between "number of poems" and "number of mentions" as measures of an author's popularity is maintained throughout the entire list of 122 authors. The correlation coefficient (which best measures this correspondence) is +0.96.

X—QUANTITATIVE STANDARDS

This study would not be complete without an attempt to determine the *amount* of memory work required by the cities whose courses of study were used. The courses varied in their types of quantative requirements. Only 34 of them contained statements which were definite enough to be used.

A few courses (most of them not among the 34 just mentioned) recognized that the same requirements should not be set up for all children. For example, the course of study for Chelsea, Massachusetts, indicated in the following words, that individual differences among pupils should be taken into consideration: "It is not expected that all children will learn equal amounts or that any will be required to master every selection, but that each teacher will use such parts as seems best suited to the taste and ability of her class and such as she can present most enjoyably and efficiently."

This sort of statement, however, was unusual. Most of the courses of study which gave definite requirements expressed them in terms of poems to be learned during the year or semester. A few stated the number of poems required each month. All such statements were changed into the number of poems per year by multiplying the number for each semester by two or the number for each month by nine.

Three courses of study (New York, Jersey City and Lexington) stated the requirements only in terms of lines of poetry. In one other case (Yonkers, N. Y.), specific poems or parts of poems were required and the lines were counted for our purpose. Where requirements were given in terms of lines per week, they were likewise converted into requirements per year, the school year being regarded as composed of 36 weeks.

The Number of Poems Required to be Memorized—Table XVIII shows the quantitative requirements of all the cities whose courses contained such requirements. These are indicated for each grade in the form of the number of poems or the number of lines to be memorized each year. Several courses contained quantitative statements for some grades, but not for all. In a few such cases the requirements were evidently unintentionally omitted, and it was sometimes possible to infer them. Whenever such an inference was made, it was indicated in Table XVIII by a foot-note.

Table XVIII gives the detailed requirements by cities. From this arrangement it is difficult to secure a total impression. Accordingly, Table XIX has been made up to show, in the form of distributions, the number of courses of study requiring in each grade the memorizing of three poems, four poems,

TABLE XVIII—MINIMUM QUANTATIVE REQUIREMENTS FOR MEMORY WORK (POETRY)

Cities	GRADES															
	I		II		III		IV		V		VI		VII		VIII	
	Poems	Lines	Poems	Lines	Poems	Lines	Poems	Lines	Poems	Lines	Poems	Lines	Poems	Lines	Poems	Lines
Austin	12[1]		12[1]		12[1]		12[1]		12[1]		12[1]		12[1]		12[1]	
Bay City	6		6		6				8		8		10		16	
Boston	9		9		9		9	100	9	150	9	150				
Chattanooga	6		6		6		6				9		9			
Cincinnati	6		6		6		6		6		6		6		6	
Denver	6		6		6		6		9[2]		9[2]		6		6	
Easton	10		8		6[2]		6		10		9		9		12	
Erie			4		4		4		4		4		4			
Fitchburg	6		6		6		6		5		5		5		5	
Flint	4[2]		4[2]		4[2]		[3]		8		8		[3]		[3]	
Fort Worth	12		[3]		11		8		5		[3]		[3]		[3]	
Harrisburg	9		9		9[5]		9[4]		9		9		9[5]		9	
Haverhill	9[4]		9[4]		9[4]		9		9[4]		9[4]		9[4]		9[4]	
Houston	7		10		7		8		8		8		7			
Indianapolis	6		6		6		6		6		6		6		6	
Jersey City		144		144		144		144		225		225		315		315
Lansing	10		10													
Lexington						216		216		216		216		216		216
Lincoln	4		4		4		4									
Newark	[3]		5		7		7		6		6		7		[3]	
New Castle	6		6		6		6		6		6		6		6	
New York City		144		144		144		144		144		144		288		288
Ogden	8		8[5]		8[5]		8[5]		8		8		20		19	
Oklahoma City	4		4		4		6		[3]		[3]		[3]			
Richmond	8		8		8		6		6		6		8		[3]	
Rochester	6		6		6		6		6		6		6		6	
Schnectady	8[2]		8		8		8		8		8		8		8	

TABLE XVIII—(*Continued*)

Cities	GRADES															
	I		II		III		IV		V		VI		VII		VIII	
	Poems	Lines	Poems	Lines	Poems	Lines	Poems	Lines	Poems	Lines	Poems	Lines	Poems	Lines	Poems	Lines
Spokane	6		6		6		6		6		6		6[5]		6	
Salt Lake City	4		4		4		[3]		[3]		[3]		[3]		[3]	
St. Paul	8		8		8		8		8		8		8		8	
Tacoma			7		10		6		6		6		6		6	
Tulsa	9		9		9		9		9		9		9		9	
Washington	9		10		10		10		8		8		8		8	
Yonkers	4	70	5	111	5	110	4	64	4	110	4	40	3	36	[3]	

(1)—Maximum

(2)—Also other materials—e.g. proverbs

(3)—No data

(4)—One long, or two short poems each month

(5)—By inference

five poems, etc. The three cases in which requirements were given only in terms of lines are necessarily omitted. This table is a condensation from a very great amount of detail. Out of the detail comes a surprising uniformity throughout the grades in respect to the number of poems required to be memorized. As far as the mode may be said to yield the best measures of general practice, we may say that the standard number of poems required in each grade, from the first through the eighth, is six. Although it probably forms no proper part of this report to discuss the suitability of the mode as an indication of the general practice, it is our belief that it affords the best measure of it in this case. This is especially true when we observe the strength of the mode. In all grades except the fifth and sixth, no other number of poems rivals six in frequency. Six poems, therefore, for each grade may be given as our best general expression of the current quantitative requirement for memory work.

As a criticism of the mode and as indicating the direction in which reasonable requirements may go, attention is called to the medians as given in Table XIX. Note that the medians are always higher than the modes. In the four upper grades, this tendency is especially marked, the differences between the median and the mode varying from 2 to 2½ poems. This indicates clearly that the distributions are skewed. If one wishes to conform to usage and if one at the same time feels like taking liberties with the mode, one is justified in doing so in the direction of greater requirements. Thus one would be conservative in requiring 6 or 7 poems in the first four grades, and 6 to 8 poems in the last four grades.

TABLE XIX—DISTRIBUTION OF THE COURSES OF STUDY ACCORDING TO THE NUMBER OF POEMS REQUIRED TO BE MEMORIZED IN EACH GRADE*

No. of Poems Required	GRADES							
	I	II	III	IV	V	VI	VII	VIII
3							1	
4	5	5	5	3	2	3	1	
5		2	1		2	1	1	1
6	9	9	10	13	8	8	7	7
7	1	1	2	1			2	
8	4	5	4	5	7	7	4	3
9	5	4	4	4	5	7	5	4
10	2	3	2	1	1		1	
11			1					
12	2	1	1	1	1		1	2
Over 12							1[1]	2[2]
Total	28	30	30	28	26	26	24	19
Mode	6	6	6	6	6	6	6	6
Median	7.0	6.9	6.9	6.8	8.1	8.1	8.0	8.5

* Table reads: 5 courses of study required a minimum of 4 poems to be memorized in the first grade; 5 required 4 poems in the second grade, etc.

(1) Twenty poems (Oklahoma City)

(2) Nineteen poems (Oklahoma City) and sixteen poems (Bay City)

XI—"MEMORY GEMS"

In the interest of completeness we have given some attention to a type of memory material which, for want of a better name, we shall call memory gems. This material consists of short quotations, mostly in verse form, which like the proverbs to be discussed in Part II are quoted in full in the courses of study. These "gems" are not presented as whole pieces. They are always brief. It is true that a few of them have real literary value, but the large majority are nothing more than jingles of questionable value. Two examples of rhymes which were found in the courses of study are given herewith. Their trivial character is evident.

"A cheerful spirit gets on quick,
  A grumbler in the mud will stick."

"Jog on, jog on, the footpath way,
  And merrily jump the stile boys,
A merry heart goes all the day,
  Your sad one tires in a mile boys."

It is ridiculous to require children to memorize such material. The teacher who desires quotations of literary value can readily find them in source books. A list of "collections of quotations" may be found in Part III of this report.

PART II—PROSE MATERIAL

Some of the courses of study required or suggested for memorizing material other than poetry. This material consisted of proverbs, biblical passages and a few longer prose selections.

I—PROVERBS

One city—Bridgeport, Conn.,—laid particular stress on proverbs. Not only were one hundred fifty of them listed to be memorized, but the reasons for doing so were carefully stated in the following words: "Proverbs are never out of place, and they have in all ages constituted a large part of the philosophy of life by which nations have lived. One of the misfortunes of our own age has been the increasing ignorance of these short, pithy sayings, which have expressed the experience of millions and helped millions more to act well their part in life."

We confess to a sense of agreement with the attitude of the writer of the Bridgeport Course of Study. It is indeed unfortunate that proverbs, in which are enshrined so much of the wisdom of the race and which have indeed played so important a part in education in the past, should be so largely neglected as these courses of study show to be the case. A proverb has been called "The wisdom of many and the wit of one." A French *bon mot* has it: "Les proverbs sont des échoes de l'expérience." Proverbs were highly regarded by the ancient Greeks,—Aristotle commending them for their use in public speaking and Quintillian for their use in writing. Erasmus, who contributed largely to the restoration of letters in Europe, devoted much time to the collection and explanation of proverbs scattered through the works of Greek and Roman writers. Moreover the study of proverbs was one of the subjects pursued in the medevial universities.

Bohn* defines a proverb as an instructive sentence, or common and pithy saying in which more is generally designed than expressed; famous for its peculiarity and elegance and therefore, adopted by the learned as well as the unlearned. Some writers regard as proverbs simple comparisons, like "As swift as Achilles," or "As cunning as a fox." Others limit proverbs to "common or pithy expressions which embody moral precepts or admitted truths." It is undoubtedly true that a great deal of insight as to the morals and standards of a people can be obtained from a study of their proverbs. Christy** uses the expression "as the people so the proverb." He also quotes the following old rhyme:

---

* Bohn, H. G. A.: Handbook of Proverbs
** Christy: Proverbs, Maxims, and Phrases of All Ages.

> "The people's voice, the voice of God we call,
> And what are proverbs but the public voice?
> Coined first and common made by common choice,
> Then sure they must have weight and truth withal."

The list of fifty proverbs given below was chosen from the entire list of 407 mentioned by the courses of study. The basis of selection was not, as in the case of the poems, frequency of mention. Indeed but few of these proverbs were mentioned in more than one course of study. The proverbs were chosen which best measured up to the following criteria:

(1.) popularity or common usage.
(2.) age—Does it occur in the standard collections of old proverbs?
(3.) value of the moral precept or advice.

There are many collections of proverbs in the English language. A great many of these simply list the proverbs and give references to earlier collections, although a few discuss them and attempt to trace their source in the Greek or Latin. Some of the collections are quite old. Most of the fifty proverbs can be found in one or more of the collections listed below. The number preceding a book will be used as the code number and placed after the proverbs which occur in that particular collection.

1. Hazlitt's "English Proverbs and Proverbial Phrases" 1834.
2. Christy's "Proverbs, Maxims, and Phrases of all Ages" 1893.
3. Marshall's "Compendions Treatise in Metre" 1554.
4. Heywood's "Proverbs" 1562.
5. Camden's "Remaines" 1614.
6. Book of Merry Riddles 1629.
7. Clark's "Paroemiologia" 1639.
8. Herbert's "Outlandish Proverbs" 1640.
9. Fuller's "Gnomologia" 1732.
10. Franklin's "Poor Richard" 1733.
11. Bland's "Proverbs," (chiefly taken from Adagia of Erasmus) 1814.

FIFTY SELECTED PROVERBS

1. A bird in the hand is worth two in the bush. (1)
2. A bird is known by its note, a man by his talk. (1)
3. A book that remains shut is but a block. (1)
4. A dwarf on a giant's shoulder sees farther of the two. (1) (8)
5. A fool and his money are soon parted. (2)
6. A lion may be beholden to a mouse. (1)
7. A penny saved is a penny earned. (1)
8. A rolling stone gathers no moss. (1) (7)
9. A still tongue makes a wise head. (1)
10. A stitch in time saves nine. (1)
11. All is not gold that glitters. (1) (4)
12. As you make your bed so you must lie in it. (1)
13. Better half a loaf than no bread. (1) (5)
14. Birds of a feather flock together. (1) (5) (6)
15. Brave actions never need a trumpet. (1)
16. Don't make a mountain of a mole-hill.
17. Every tub must stand upon its bottom. (1)
18. Experience keeps a dear school; but fools learn in no other. (1)
19. Fair words hurt not the mouth. (1) (5)
20. Fine feathers do not make fine birds. (1)
21. Genius begins great works, labor alone finishes them.
22. God helps them that help themselves. (1) (10)
23. Good health is above wealth. (1)
24. Honesty is the best policy. (1)
25. If you wish a thing done, go; if not, send. (1)
26. Imitation is the sincerest flattery. (1)
27. It is hard for an empty bag to stand upright. (1)
28. Keep thy shop and thy shop will keep thee. (1) (10)
29. Knowledge is power. (1)
30. Labor conquers all things. (Labor omnia vincit.)
31. Little strokes fell great oaks. (1)
32. Little things are great to little men.
33. Lost time is never found again. (1)
34. Many hands make light work. (1) (4)
35. Never count your chickens before they are hatched.
36. One today is worth two tomorrows. (1)
37. Practice makes perfect. (1)
38. Procrastination is the thief of time. (1)

39. Rome was not built in a day. (1) (4)
40. Second thoughts are best. (1)
41. The army that comes off best loses some. (1)
42. The burnt child dreads the fire. (1)
43. The cat would eat fish, but would not wet her feet. (1) (4)
44. The mouse that hath but one hole is easily taken. (1)
45. The proof of the pudding is the eating. (1) (4) (7)
46. There is no royal road to learning. (1)
47. Waste not, want not. (1)
48. What cannot be cured must be endured. (1)
49. Where there's a will there's a way. (1)
50. While there is life there is hope.

II—Biblical Material

The passages from the Bible consisted of psalms and selections, such as the Ten Commandments and the Beatitudes, as well as of short quotations. No argument is needed to convince one of the ignorance of the present generation regarding the Bible. In a literature class of the present day the simplest Biblical allusion is likely to be unfamiliar to a majority of the students. Since children do not learn the Bible in the home, undoubtedly it is the duty of the school to teach at least those portions of it which have particular literary and moral value. This is quite apart from the use of the Bible in religious instruction. It has great intrinsic literary value, and its influence upon other literature—both as to thought and form—has been even greater.

Of all Biblical material the twenty-third psalm was most frequently mentioned by the courses of study. It was listed 21 times. The nineteenth psalm was second with a total frequency of 17. It is evident that these psalms are as frequently required for memory as some of the most popular poems. Psalms I and XXIV were each mentioned 9 times; Psalms C and CXXI 7 times; and the Beatitudes 5 times. The rest of the longer Biblical passages fell below 5.

The frequency with which verses and other short quotations were mentioned was of little significance. Some of them occurred in the lists of proverbs, and many were not mentioned by more than one course of study.

The entire list of Biblical material is given below. It consists of whole selections and short quotations. In all cases the complete Biblical reference has been supplied. Not infrequently the quotations occur in more than one passage in the Bible.

No attempt has been made to show all the places where the quotations may be found.

BIBLICAL MATERIAL

A.—WHOLE SELECTIONS

		Times Mentioned
1.	Exodus 20:2-17 (Ten Commandments)	3
2.	Psalms 1	9
3.	“ 8	1
4.	“ 19	17
5.	“ 23	21
6.	“ 24	9
7.	“ 37	1
8.	“ 42	1
9.	“ 46	3
10.	“ 65	1
11.	“ 67	2
12.	“ 90	4
13.	“ 91	1
14.	“ 95	1
15.	“ 96	1
16.	“ 100	7
17.	“ 103	2
18.	“ 121	7
19.	“ 148	2
20.	Poverbs Ch. 1	1
21.	“ “ 10	1
22.	“ “ 15	1
23.	“ “ 20	1
24.	“ “ 22	1
25.	Matt. Ch. 5 (The Beatitudes)	5

B.—SHORT QUOTATIONS

Quotation	Reference	Times Mentioned
1. A living dog is better than a dead lion	Eccl. 8:4	1
2. A merry heart doeth good like a medicine	Prov. 17:22	2
3. A soft answer turneth away wrath	Prov. 15:1	3
4. A wise son maketh a glad father	(Prov. 10:1 (Prov. 15:20	1
5. A word fitly spoken is like apples of gold in pictures of silver	Prov. 25:11	1
6. And with what measure ye mete, it shall be measured to you again	(Matt. 7:2 (Mark 4:24 (Luke 6:38	1
7. As he thinketh in his heart, so is he	Prov. 23:7	1
8. Be not deceived	(1 Cor. 6:9 (1 Cor. 15:33	1
9. Be not wise in your own conceits	Romans 12:16	9

10. Blessed are the merciful for they shall obtain mercy	Matt. 5:7	1
11. Blessed is the man that endureth temptation	James 1:12	1
12. Cast thy bread upon the waters; for thou shalt find it after many days	Eccl. 11:1	1
13. Enter into His gates with thanksgiving and into His courts with praise. Be thankful unto Him and bless His name.	Psalm 100:4	1
14. Evil communications corrupt good manners	1 Cor. 15:33	1
15. Get wisdom	Prov. 4:5	1
16. He that is slow to anger is better than the mighty and he that ruleth his spirit than he that taketh a city	Prov. 16:32	1
17. If the blind lead the blind, both shall fall into the ditch.	Matt. 15:14	1
18. In the sweat of thy face shalt thou eat bread	Gen. 3:19	1
19. It is more blessed to give than to receive	Acts 20:35	2
20. Judge not that you be not judged	Matt. 7:1	8
21. Keep thy tongue from evil, and thy lips from speaking guile	Psalm 34:13	1
22. Let another man praise thee and not thine own mouth	Prov. 27:2	1
23. Let the wicked forsake his way, and the unrighteous man his thought	Is. 55:7	1
24. Man doth not live by bread alone	Deut. 8:3	1
25. Mark the perfect man and behold the upright, for the end of that man is peace	Psalm 37:37	1
26. Pride goeth before destruction and a haughty spirit before a fall	Prov. 16:18	1
27. Prove all things; hold fast that which is good	1 Thess. 5:21	1
28. Remember Now Thy Creator	Eccl. 12:1	1
29. Render therefore to all their dues	Rom. 13:7	1
30. Righteousness exalteth a nation	Prov. 14:34	1
31. The borrower is servant to the lender	Prov. 22:7	1
32. The fool hath said in his heart, there is no God	(Psalm 14:1 Psalm 53:1	1
33. The truth shall make you free	John 8:32	1
34. Therefore, all things whatsoever that men should do to you, de ye even so to them	Matt. 7:12	10
35. Thou shalt love thy neighbor as thyself	(Rom. 13:9 (Mark 12:31 (Jam. 2:8 (Gal. 5:14	1
36. Unstable as water thou shalt not excel	Gen. 49:4	1
37. Unto the pure all things are pure	Titus 1:15	1
38. Whatsoever a man soweth, that shall he also reap	Gal. 6:7	10
39. Whatsoever things are true, whatsoever things are honest, whatsoever things are just, whatsoever things are pure, whatsoever things are lovely, whatsoever things are of good report, if there be any virtue and if there be any praise, think on these things	Phil. 4:8	2
40. Whatsoever thy hand findeth to do, do it with thy might	Eccl. 9:10	1
41. Woe unto them that call evil good, and good evil	Is. 5:20	1

III—LONGER PROSE SELECTIONS

The prose selection which was most frequently mentioned was Lincoln's "Gettysburg Address." This selection was as frequently required for memory as many of the most popular poems. The grade to which it was almost universally assigned was the eighth. The chances are that most children who graduate from the elementary school will have been required to memorize this selection. The other prose selections which occurred in the courses of study were practically limited to patriotic material. None of these received a high frequency of mention. In fact many of them occurred but once. The following list is representative of the prose material.

1. McDowell—Civic Creed.
2. Patrick Henry—An Appeal to Arms.
3. Wm. T. Page—The American Creed.
4. Pledge of Allegiance.
5. Preamble to the Constitution.
6. First paragraph of the Declaration of Independence.
7. The Athenian Oath.

This material is undoubtedly valuable for memory work. But the body of prose material valuable for this purpose is very extensive, and these few selections can hardly be said to represent it. At least two selections—The Preamble to the Constitution and the first paragraph of The Declaration of Independence are often memorized in connection with the work in history and civics. The investigation shows clearly that the prose material, as used in memory work, is limited in amount and restricted in character.

PART III—BIBLIOGRAPHIES

I—COURSES OF STUDY

The following are the fifty courses of study which were used in this report.

Allentown, Pennsylvania—Course of Study in English, Grades One to Eight, Public Schools of Allentown, Pennsylvania, 1918.

Altoona, Pennsylvania—Stories and Poems for Each Grade, School District, Altoona, Pennsylvania.

Austin, Texas—Course of Study of the Austin Public Schools for the Primary and Intermediate Grades for the Scholastic Year, 1916.

Bay City, Michigan—Manual and Course of Study of the Bay City Public Schools, Bay City, Michigan, 1918.

Berkeley, California—A Course of Study for the Elementary Schools of Berkeley, California, 1917-1918.

Boston, Massachusetts—A Provisional Course of Study for the Elementary Schools. (School Document No. 8, Boston Public Schools.)

——————— Syllabus for the Elementary Schools, (School Documents Nos. 11, 15, 16, 1915; No. 11, 1916; Nos. 18, 19, 1917; Boston Public Schools.)

Bridgeport, Connecticut—Syllabus of English, Bridgeport, Connecticut, 1916.

Brookline, Massachusetts—An Outline of Work in Literature for the Primary and Grammar Schools of Brookline, Massachusetts, 1915.

Chattanooga, Tennessee—Course of Study, Public Schools, Chattanooga, Tennessee, 1918.

Chelsea, Massachusetts—Provisional Courses of Study for the Public Elementary and Junior High Schools, of the City of Chelsea, Massachusetts, 1917.

Cincinnati, Ohio—Tentative Course of Study in English: Reading, Language, Spelling, Gades I-VIII, Cincinnati Public Schools, Cincinnati, 1918.

Denver, Colorado—Course of Study for English, Grades One to Eight Inclusive, Denver Public Schools, 1918.

Easton, Pennsylvania—Outline of Course of Study, Easton School District, 1918-1919.

Erie, Pennsylvania—Course of Study in Reading, Language, Spelling, Penmanship, Arithmetic, History, Civics,

Hygiene, and Music, Erie Public Schools, 1918.

Fitchburg, Massachusetts—Elementary Course of Study for the Public Schools, Fitchburg, Massachusetts, 1919.

Flint, Michigan—Course of Study, Books I-II, Flint Public Schools, 1918.

Fort Worth, Texas—Outline of Course of Study of Fort Worth Public Schools, 1918-1919.

Harrisburg, Pennsylvania—Outline of Course of Study, Harrisburg Public Schools, 1917.

Haverhill, Massachussets—Course of Study for the Public Schools, Part I, Haverhill, Massachusetts, 1919.

Houston, Texas—Course of Study of the Public Schools of the Independent School District of the City of Houston, Texas, 1917-1918.

Indianapolis, Indiana—Course of Study in English in the Elementary Schools with Suggestions and Directions to Teachers, Indianapolis Public Schools, 1914.

Jersey City, New Jersey—Course of Study and Manual of Instruction for Elementary Schools, Jersey City, New Jersey, 1917.

Kansas City, Missouri—Tentative Course of Study for the Elementary Schools: Reading, Language, Spelling, Kansas City, Missouri Public Schools, 1915.

Lansing, Michigan—Course of Study, Lansing Public Schools, 1916.

Lexington, Kentucky—Course of Study of the Elementary, Junior High, and Senior High Schools of Lexington, Kentucky.

Lincoln, Nebraska—Supplement to the Course of Study in English in the Elementary, Junior High and Pre-vocational Schools in Lincoln, Nebraska, 1917.

Louisville, Kentucky—Language Outline with Type Lessons, Louisville, Kentucky, 1915.

Madison, Wisconsin—Course of Study, Madison, Wisconsin, 1912.

Nashua, New Hampshire—Course of Study for the Nashua Public Schools, 1913.

Newark, Ohio—Course of Study and Manual of Instruction for the Public Schools of Newark, Ohio, 1915.

New Castle, Pennsylvania—Outline of Instruction in the Elementary Schools of New Castle, Pennsylvania, 1916.

New York City, New York—Courses of Study and Syllabuses in Ethics, English, Geography, Penmanship, for the Elementary Schools of the City of New York, 1917.

Ogden, Utah—Syllabi of Courses of Study with a Brief Statement of the Rules and Regulations Governing the Schools, Ogden, Utah, 1917.

Oklahoma City, Oklahoma—Course of Study and Manual for the Primary and Grammar Grades, Oklahoma City Schools, 1917.

Oshkosh, Wisconsin—Manual of the Public Schools of Oshkosh, Wisconsin, 1915-1916.

Pasadena, California—Poems for Use in the Elementary Schools, compiled by Jeremiah Rhodes, Pasadena, California, 1918.

———————— —Words from the Masters, compiled and arranged by Jeremiah Rhodes, Pasadena, California.

Pittsfield, Massachusetts — Language Outline, Public Schools, Pittsfield, Massachusetts, 1913.

Portland, Oregon—Elementary Course of Study, Portland, Oregon, 1918.

Richmond, Virginia—Course of Study in English including Reading, Phonics, Literature, Language, Grammar, Spelling, Penmanship for Elementary and Junior High Schools, Richmond, Virginia, 1918.

Roanoke, Virginia—Course of Study for the Primary Schools, Roanoke City Schools, Roanoke, Virginia.

Rochester, New York—Outline of the Course of Study of the Public Schools of the City of Rochester, New York: English, 1917.

St. Paul, Minnesota—Provisional Course of Study in Poems, St. Paul, Minnesota, 1917-1918.

Salt Lake City, Utah—Courses of Study, Public Schools of Salt Lake City, Utah, 1917.

Schenectady, New York—Course of Study for the Elementary and Evening Schools, Schenectady, New York, 1914.

Scranton, Pennsylvania—Course of Study for the Elementary Schools, Scranton, Pennsylvania, 1915.

Spokane, Washington—Course of Study, Reading, Elementary Public Schools, Spokane, Washington, 1918.

Tacoma, Washington—Course of Study for the Elementary Schools of Tacoma, Washington, 1917.

Tulsa, Oklahoma—Course of Study, Tulsa City Schools, 1918-1919.

Washington, D. C.—Course of Study in English, Elementary Schools, Washington, D. C., 1914.

Yonkers, New York—Annual Report of the Superintendent of Schools, Yonkers, New York, 1915-1916.

II—COLLECTIONS OF POETRY

1. ————————. *One Hundred and One Famous Poems.* Chicago: Cable Piano Company.
2. Beeson, R. Katherine. *Child's Calendar Beautiful, The.* New York: Chas. Scribner's Sons.
3. Bellamy, B. W. and M. W. Goodwin. *Open Sesame.* New York: Ginn & Co.
4. Browne, Francis F. *Golden Poems by British and American Authors.* Chicago: A. C. McClurg Co.
5. Bryant, William C. *Complete Works.* New York: D. Appleton & Co.
6. Burt, Mary E. *Poems Every Child Should Know.* New York: Crosset & Dunlap.
7. Campbell, Loomis J. *New Franklin Fifth Reader.* New York: American Book Company.
8. Cary, Alice and Phoebe. *Complete Works.* New York: Houghton Mifflin Co.
9. Chisholm, Lowry. *Golden Staircase: Poems and Verses for Children.* New York: G. P. Putnam & Sons.
10. Emerson, Ralph, W. *Complete Works.* New York: Houghton Mifflin Co.
11. Field, Eugene. *Complete Works.* New York: Chas. Scribner's Sons.
12. Granger, Edith. *Index to Poetry and Recitation.* Chicago: A. C. McClurg & Co.
13. Hazard, Bertha. *Three Years with the Poets.* New York: Houghton Mifflin Co.
14. Holmes, Oliver W. *Complete Works.* New York: Houghton Mifflin Co.
15. Jones, Lottie E. and S. W. Black. *New Era Third Reader.* Chicago: Eaton & Co.
16. Larcom, Lucy. *Complete Works.* New York: Houghton Mifflin Co.

17. Lear, Edward. *Nonsense Books.* Boston: Little Brown & Co.
18. Longfellow, Henry W. *Complete Works.* New York: Houghton Mifflin Co.
19. O'Neill, A. T. L. *Recitations for Assembly and Classroom.* New York: The Macmillan Co.
20. Rice, Wallace and Frances. *The Humbler Poets.* Chicago: A. C. McClurg Co.
21. Riley, James Whitcomb. *Harper Edition.* New York: Harper & Bros.
22. Scott, Sir Walter. *Lady of the Lake.* New York: C. E. Merrill.
23. Sherman, F. D. *Little Folk Lyrics.* New York: Houghton Mifflin Co.
24. Stedman, Edmund C. *American Anthology.* New York: Houghton Mifflin Co.
25. Stevenson, Burton E. *Home Book of Verse.* New York: Henry Holt & Co.
26. Stevenson, Burton E. *Home Book of Verse for Young Folks.* New York: Henry Holt & Co.
27. Stevenson, Robert L. *A Child's Garden of Verses.* New York: G. P. Putnam & Sons.
28. Tennyson, Alfred, Lord. *Complete Works.* New York: The Macmillan Co.
29. Whittier, John G. *Complete Works.* New York: Houghton Mifflin Co.
30. Wiggin, Kate Douglas and Smith, Nora A. *Golden Numbers: Poems for Children and Young People.* New York: Houghton Mifflin Co.
31. Wiggin, Kate Douglas and Smith, Nora A. *Pinafore Palace: Book of Rhymes for the Nursery.* New York: Crosset & Dunlap.
32. Wiggin, Kate Douglas and Smith, Nora A. *Posy Ring: Verses and Poems for the Youngest Children.* New York: Houghton Mifflin Co.

III—COLLECTIONS OF QUOTATIONS

1. Bartlett, John. *Familiar Quotations.* Boston: Little, Brown & Co. 1914.
2. Dalbiac, Philip H. *Dictionary of Quotations.* London: Swan Sonnenschien & Co. 1896.

3. Douglas, Charles Noel. *Forty Thousand Quotations, Prose and Poetical.* New York: Sully & Kleinteich. 1916.
4. Walsh, Wm. S. *The International Encyclopedia of Prose and Poetical Quotations.* Philadelphia: The J. C. Winston Co. 1908.

IV—READERS

1. Baker, Franklin and Thorndike, Ashley. *Everyday Classics.* Third to Eighth Readers. New York: The Macmillan Co. 1917-1918.
2. Bolenius, Emma Miller. *The Boys' and Girls' Readers.* Fourth to Sixth Readers. Boston: Houghton Mifflin Co. 1919.
3. Briggs, Thomas H. *Reading-Literature.* Seventh Reader. Chicago: Row Peterson & Co. 1915.
4. Burchill, Georgine, Ettinger, Wm. L., and Shimer, Edgar D. *The Progressive Road to Learning.* Books One to Six. Silver Burdett & Co. 1909-1919.
5. Coe, Ida and Christie, Alice J. *Story Hour Readers.* Books Two and Three. New York: American Book Co. 1914.
6. Dyer, Franklin B. and Brady, Mary J. *The Merrill Readers.* Seventh and Eighth Readers. New York: Charles E. Merrill Co. 1916.
7. Elson, Wm. H. and Keck, Christian. *Elson Primary School Reader.* Books Two to Four. *Elson Grammar School Reader.* Books One to Four. Chicago: Scott, Foresman & Co. 1909-1912.
8. Fassett, James H. *Beacon Readers.* First to Fifth Readers. Boston: Ginn & Co. 1913-1918.
9. Free, Margaret and Treadwell, Harriette T. *Reading-Literature.* First to Sixth Reader. Chicago: Row Peterson & Co. 1911-1915.
10. Gordon, Emma K. *The Gordon Readers—New Series.* First to Fourth Readers. Boston: D. C. Heath & Co. 1918.
11. Hervey, Walter L. and Hix, Melvin. *The Horace Mann Readers.* Second to Eighth Readers. New York: Longmans, Green & Co. 1909-1918.

12. Horn, Paul W., Shurter, Edwin D., and Baugh, Lila. *New American Readers.* Books One to Five. Boston: Ginn & Co. 1919.
13. Hyer, Frank S. *The Laurel Readers.* First and Second Readers. Chicago: Laurel Book Co. 1917.
14. Kendall, Calvin N., Stevens, Marvin P., and Townsend, Caroline L. *The Kendall Series of Readers.* First to Fourth Readers. Boston: D. C. Heath & Co. 1918.
15. Laselle, Mary A. *The Home and Country Readers.* Books One to Four. Boston: Little Brown & Co. 1918.
16. Lewis, Horner P. and Lewis, Elizabeth. *Lippincott's Readers.* Third to Fifth Readers. Philadelphia: J. B. Lippincott Co. 1910-1913.
17. Metcalf, John C., Withers, Sarah, and Browne, Hetty S. *The Literary World.* Sixth to Eighth Readers. Atlanta: B. F. Johnson Publishing Co. 1919.
18. Murray, Clara. *The Wide Awake Readers.* Second to Fourth Readers. Boston: Little Brown & Co. 1908-1917.
19. Spaulding, Frank E. and Bryce, Catherine T. *Aldine Readers.* Books One to Four. New York: Newson & Co. 1916-1919.
20. Summers, Maud. *The Summers Readers.* First to Third Readers. New York: Lloyd Adams Noble. 1909-1914.
21. Tate, W. K., Withers, Sarah, and Browne, Hetty S. *The Child's World.* First to Fifth Reader. Atlanta: B. F. Johnson Publishing Co. 1917.
22. Van Sickle, James H. and Seegmiller, Wilhelmina. *The Riverside Readers.* Second to Eighth Readers. Boston: Houghton Mifflin Co. 1911-1913.
23. Young, Ella Flagg and Field, Walter Taylor. *The Young and Field Literary Readers.* Books One to Six. Boston: Ginn & Co. 1914-1916.

BULLETINS OF THE BUREAU OF EDUCATIONAL RESEARCH

B. R. BUCKINGHAM, *Editor*

BUREAU OF EDUCATIONAL RESEARCH—BULLETIN NO. 4

3

MENTAL TESTS FOR SCHOOL USE

BY

CHARLES E. HOLLEY, Ph.D.

Formerly Assistant Director
Bureau of Educational Research
University of Illinois

Price 50 Cents

PUBLISHED BY THE UNIVERSITY OF ILLINOIS
URBANA

COPYRIGHT, 1920
BY THE UNIVERSITY OF ILLINOIS

EDITORIAL INTRODUCTION

In the spring of 1919 Doctor Holley, while acting as assistant director of the Bureau of Educational Research of the University of Illinois, tried out six group intelligence scales in the schools of Champaign, Illinois. A kind of survey, narrow but intensive in character, was thus afforded. The data, however, with a little more analysis could be made to yield important results as to the reliabilty and validity of each of these six scales as instruments for measuring intelligence. With this thought in mind Doctor Holley carried out some of the necessary analyses and wrote the monograph which follows.

Of the six tests, three have become popular in a large way. They are the "Otis Group Intelligence Scale," the "Primer Scale," and the "Virginia Delta I" (now known as the "Intelligence Examination, Delta 2"). Besides the six which were used in this investigation there were at least three others which might have been used. In all there appear to have been nine rather well-known tests at the time the survey at Champaign was started.

Since then the number has been materially increased. Not only did several new tests come out during the school year, 1919-1920, but at least three scales, complete in every essential detail, have been published this summer in anticipation of the "fall trade." The World Book Company announces Terman's "Group Test of Mental Ability"; Lippincott annouces the "Dearborn Group Tests of Intelligence"; and the Bureau of Educational Research of the University of Illinois announces the "Illinois General Intelligence Scale." It is apparent that the movement to measure intelligence by means of group tests is well under way.

Under these circumstances school people are inquiring somewhat anxiously, "Which among all the intelligence tests is best?" Like most general questions, this has no general answer. The "best" test is the one which is most appropriate. It may not be best at all times, with all pupils, and for all purposes. The term "best" therefore needs qualification.

Nevertheless, no matter what the qualifications, there are certain characteristics which a good test—to say nothing of the best one—should possess. It should not require too much time to administer. It should be capable of rapid and objective rating. It should correlate highly, but not too highly, with teachers' estimates of scholarship—say about 0.60. It should discriminate unmistakably between levels of intelligence which are known on other grounds to be different—e.g., the levels at different ages or grades. The subordinate exercises of which it is composed should test important mental traits and should contribute to the total score amounts proportional to the importance of these traits. Scores in the subordinate exercises should be relatively independent, for otherwise they merely tell the same story. Moreover, like the scales of which they form a part, they should discriminate between levels known to be different. Both the entire scale and its subordinate exercises should yield very few zero scores and very few scores of the highest possible value. Indeed, there should not be many scores even in the region of these extremes.

All of the scales in this investigation were examined with reference to these points. The method is of necessity largely statistical; but the outcome is practical enough. Certain very definite recommendations and suggestions are made. It is believed, therefore, that with reference to a few important tests the serious student when asking the question, "Which test is best," will find, if not a general answer, at least something fundamental and satisfactory.

B. R. BUCKINGHAM,

Director, Bureau of Educational Research,
University of Illinois.

August 26, 1920.

9

TABLE OF CONTENTS

PART I—THE PRESENT STATUS OF MENTAL TESTING

PAGE
A. Uses of Mental Tests 9
B. What Mental Tests Measure 16
C. Types of Mental Tests 20

PART II—COMPARISON OF GROUP MENTAL SCALES

A. Introduction 25
B. Administration of Tests 25
C. The Time Element 35
D. Comparisons of Total Scores 36
E. Analysis of the Individual Tests in the Intelligence Scales 47
F. General Comparison of Scales 62
G. Summary and Comments 65

PART III—MENTAL SURVEY OF THE CHAMPAIGN PUBLIC SCHOOLS

A. Introduction 66
B. Variations between Schools and Grades 66
C. Summary and Comments 84
Appendix . . . . , ... , . . 86

INDEX TO TABLES

NUMBER PAGE

I Total Scores in the Otis Group Intelligence Scale 26
II Total Scores in the Theisen-Flemming Classification Test 27
III Distribution of Scores in Pressey's Primer Scale 28
IV Scores in Virginia Delta I. 29
V Distribution of Scores in Sentence Vocabulary Scale 31
VI Pupils Marking Vocabulary Sentences Correctly 33
VII Time Consumed in Administering Scales 35
VIII Rate of Scoring by Clerical Workers 36
IX Correlations between Intelligence Scores and Teachers' Ratings 37
X Teachers' Scholarship Ratings. First Grade 38
XI Teachers' Scholarship Ratings. Fourth Grade 38
XII Distribution of Coefficients of Correlation between Intelligence Scores and Scholarship Ratings 39
XIII Correlations between Scales 41
XIV Distribution of Inter-Scale Correlations 42
XV Differences between Grade Averages with Reliabilities for Scores in Each Scale 44
XVI Coefficient of Variability of Differences between Scores for Successive Grades 46
XVII Discrimination between Larger Grade Ranges for Intelligence Scales 48
XVIII Discriminative Power of Otis Scale 51
XIX Discriminative Power of Classification Test 52
XX Coefficients of Variability of Individual Tests for Otis Scale 54
XXI Coefficients of Variability of Individual Tests for Classification Test 54
XXII Coefficients of Variability of Individual Tests for Virginia Delta I. 55
XXIII Coefficients of Variability of Individual Tests for Whipple's Group Test 55
XXIV Coefficients of Variability of Individual Tests for Primer Scale 56
XXV Coefficients of Variability for Vocabulary Scale 56
XXVI Correlation between Equivalent Tests 58
XXVII Number of Repetitions of a Test needed to Secure Reliability Indicated 60
XXVIII Percent of Zero Scores for Different Scales 61
XXIX General Comparisons of the Six Scales 64
XXX School Scores for Each Scale by Grades 68
XXXI Scores in School No. 9 for Each Scale by Rooms 69
XXXII Medians and Quartile Deviations 70
XXXIII Distribution of Pupils according to Their Dispersion from the Median 71
XXXIV Distribution of Pupils in School No. 9 according to Their Dispersion from the Median 74
XXXV Standard Scores for Indicated Ages 76
XXXVI Distribution of Intelligence Quotients by Schools and Grades 77
XXXVII Distribution of Intelligence Quotients in the Eighth Grade of School No. 9 79
XXXVIII Distribution of Intelligence Quotients in the High School 79
XXXIX Implications of Intelligence Quotients 80
XL Intelligence Quotient Distributions for the Entire School 81
XLI Age-Intelligence-Quotient Distribution for Sixth Grade 83
XLII Picture Completion Test 89

MENTAL TESTS FOR SCHOOL USE

PART I—THE PRESENT STATUS OF MENTAL TESTING

A—Uses of Mental Tests

I. *The Recognition of Feeble-minded Children*—Mental ability differs among individuals from very superior to very inferior with every possible gradation between these extremes. Degrees of difference in intelligence are usually unnoticed in a community because routine life does not reveal them to the general observer. We note only the grosser variations and these, as a rule, mainly when individuals are markedly defective. The village simpleton is a familiar figure. He is the roustabout who does light chores. He is the object of ridicule, the butt of the jokes of his more intelligent associates. Among the great mass of humanity, however, discriminations are not made; and when the question of competence is raised in the school or elsewhere, there is no satisfactory basis on which agreement may rest.

In reality feeble-mindedness is present in nearly every community to a greater degree than has ever been recognized. The work of the Psychological Service of the United States Army leads one to believe that the number of people who, when mature, do not exceed the mental development of the average nine-year-old child, is probably two or three out of each hundred of the population. Some authorities have placed the dividing line between feeble-mindedness and normality between the ages of ten and eleven or eleven and twelve. On this basis the feeble-minded would probably comprise from 5 to 10 percent of the the total population of the United States.

Feeble-minded people are sadly limited in their ability to adjust themselves to social conditions. They are weak in their control of mind and body and difficult to teach. They have very poor memories and very poor discriminative power. Constant repetitions are required in order to teach them the simplest things. Many of them never learn to read or to spell or to do simple arithmetic problems even if they are kept in school for the entire compulsory education period. This applies especially to those whose mental ability does not exceed that of the average seven-year-old child. Those whose mental development is a little better than this may do something with the ordinary schoolroom tasks but the results are hardly proportionate to the cost in time and effort. The best that can be done for these people is to train them along manual lines. They can often be taught to do the ordinary home tasks of sweeping, dusting, washing dishes, peeling potatoes, bringing in coal and wood, mowing the lawn, chopping wood, and running errands of a simple nature. Even with constant supervision, there is little hope that those whose mental rating is below seven years can be made self-supporting. Many men of eight- and nine-year mental ability, however, are getting along in the industrial world at the ordinary tasks which employ unskilled labor. They make a poor living to be sure, but they eke out an existence.

The lower types of the feeble-minded have so little mental ability that they seldom engage in crime. Occasionally feeble-minded women even of low grade become social menaces but they do not usually take the aggressive part in their misdemeanors. The higher grades of the feeble-minded, however, are a real social problem, for they are capable of participating in crimes of various sorts. Although comparatively few crimes are committed by real mental defectives, criminals actually exhibit every level of intelligence. Indeed they are more often characterized by moral than intellectual abnormality. On the other hand, people who are subnormal mentally are often model citizens, when social conduct is considered, because they have been trained to live correctly. Many criminals are defective in intelligence, but not all mental defectives are criminals even potentially.

Mental tests are of value in detecting more accurately than personal judgment the different grades of feeble-mindedness. The school may use the results of these tests in determining those for whom ordinary school work is entirely unsuited. These pupils should be given school tasks, as far as possible, that are of the manual type, because this is for them the most hopeful field of training. Even in this work the same returns should not be expected that would be secured from normal children. It is wasteful to spend a markedly disproportionate amount of the school funds on this part of the population, though a portion of the expense may be justified on the ground that the normal children profit by the segregation of the defectives.

II. *The Recognition of Mentally Backward Children*—Above the feeble-minded in mental development come those whom we call the mentally backward. These comprise from 10 to 20 percent of the population, depending upon the criteria that are set up as the dividing lines between feeble-mindedness, backwardness, and normal development. This is the class of our population from which, as a rule, petty criminals come. These are the people who are decidedly maladjusted under present conditions and who populate our slums and hovels.

The backward learn slowly at school. They have poor memories, poor discriminative powers, and mediocre reasoning ability. If they are to be taught anything the process involves a large number of repetitions. As a rule, even when they have reached their physical maturity, they are still like children in many respects. They live in the present and care little and provide little for the future. In the schoolroom they are usually retarded; but they may have enough ability to do, in a mediocre way, the work of the grade in which they are classified if they are given extra attention. Teachers often fail to appreciate the difference between their chronological ages and those of their classmates, and, hence they fail also to detect their backwardness. Developed along some lines these backward children have instincts and emotional reactions which are those of children of their own age. This side of their nature enables them at times to surprise the teacher with what seem to be bright responses; and for this reason they are often rated higher in intelligence than they should be.

The backward children in our schools should have special treatment. If put in classes by themselves they can be given the requisite repetitions of subject-matter; and they may thus learn at the rate of which they are capable. They need a special course of study built for their needs. When in the classroom with normal children they are continuously required to do things more quickly than their mental ability permits. As a consequence they fail, although if they were given more time they could succeed. They acquire the habit of failure, of which so much has been written. Mental tests would reveal the true situation and permit proper provisions to be made.

III. *The Recognition of Normal Children who are Apparently Abnormal*—Mental tests may also be used to detect normal children who are not making the progress of which they are capable. It occasionally happens that children who have average ability fail to keep up in their school work. In such cases it would be very profitable for the teacher to take the extra time needed to coach these children in order that they may progress normally. Special attention given to normal children who have "lost out" for some reason or other, often pays very well—a point recognized by those superintendents who have organized "opportunity classes" to provide for them. Teachers of such classes, however, sometimes waste their energies on really defective children because the normal children have not been differentiated from them. If mental tests are to be used for this purpose they should be given along with tests in school subjects and, if a child is mentally normal according to the mental tests and retarded when judged by the school tests, it is obvious that extra attention given to his weakness will help to eliminate or at least to lessen it. Normal children are sometimes temperamental and fail to progress because they get "at outs" with the teacher. Situations such as these may be revealed readily and the proper remedies may then be applied.

IV. *The Discovery of Superior Children*—Mental tests have special value in the selection of superior children for special classes. These children may be just as much above the average as the backward and feeble-minded are below it. Many of them could do the ordinary work of the eight years in the elementary schools in one, two, or three years less time. There are two ways in which provision may be made for these children. One is by allowing them to skip grades now and then. This device is not to be recommended without qualifications. If a child skips a grade he misses some of the vital things at times and may be handicapped in this way. Often, however, it may be better for the markedly superior child to skip grades and thus reach his school level, than to move along in lockstep style.

The other way is to provide special classes for superior children. If a number of these children are detected by the administration of mental tests, they may be placed in a class by themselves. Under these conditions they will make much more rapid progress than they would in regular classes. Such a special class makes unusual demands on the teacher, and great care must, therefore, be taken in selecting the one who is to lead a group of superior children. The teacher must be above the average in scholarship and be able at all times to keep up with the children in their thinking. In these special classes the work may be adjusted to the needs of the children. In some subjects they may be able to progress more rapidly than the average child. In others, the emphasis probably should be placed on supplementing the work, thus making it richer in content.

If superior children are kept in classes with normal children they often acquire bad habits. They are kept marking time at a point far below their possible working efficiency and, thus, acquire all the undesirable characteristics of mischievous children. Habits of idleness, disorder, and general inefficiency are often the result of this maladjustment. On the other hand, superior children should not be pushed too rapidly in school work as is often done when they receive extra promotions. If they are sent along at too rapid a pace they often reach levels where their mental ability is not equal to some of the tasks which are set before them. This is because the subject-matter has been graded to meet the needs of normal children whose emotional lives have matured in a definite relationship with their mental lives. Superior children with their unusual mental development are often merely normal in their emotional lives, having for example fourteen-year-old minds in ten-year-old bodies with ten-year-old emotions. When a superior child is expected to feel and think in the same terms as a child several years his chronological senior he is often unable to do so. This situation implies that if special classes become common, it may be necessary to modify the subject-matter used in classes for the gifted so that it may be fitted to them.

One argument often made against the rapid promotion of superior children is that they are soon thrown into companionship with older children. This criticism is a serious one. It is, however, anticipated by the provision for special classes advocated—a provision which groups a number of these superior children together. Where it is impossible to form classes for superior children, as will usually be the case in small school systems, one should consider the situation carefully before making extra promotions. Yet a superior child will sometimes reach a place where there is almost nothing for him to do in the grade in which he is placed. A child without something to do is a menace to himself. Under these circumstances, it may be the plain duty of the school to promote him.

V. *The Grading of Children for General Promotion*—Another use that may be made of mental tests is to reveal individual differences as a basis for grading and promotion. It is often necessary to regrade children who are changing schools. Under these conditions one cannot rely upon their marks because, coming from different schools and from different teachers, the children have been rated according to different standards. A good mental test will enable one to regrade the children in a fairly satisfactory way. These classifications can then be compared with the scholarship achievements of the pupils during the first month of the year and minor adjustments may be made. When pupils pass from one type of school to another—as from elementary to high school, or from high school to college—the application of mental tests as a basis of judging fitness for entrance and of sectioning is important. It is probable that the near future will see an extensive use of mental tests as a means of determining fitness to enter new schools.

The use of mental tests is particularly appropriate in the junior high school where sections are often formed on the basis of mental ability. In systems where a grade has four or five sections organized on this basis, it has been found that the best sections often do twice as much work as the poorest.

VI. *The Determination of General and Special Ability for Educational and Vocational Guidance*—The near future will also probably witness the extended use of mental tests in another field—the field of vocational and educational guidance. Under present conditions there are few tests which can be recommended even in a limited way as suitable for this work. This situation, however, is likely to be temporary. We are making rapid strides in the preparation of mental tests. It is probable that the year 1920 will see the publication and the standardization of a number of mental tests both general and special. Some of these, no doubt, will be suitable for this work.[1]

There are two phases of this problem. It may be attacked from the point of view of so-called general intelligence. A certain degree of mental ability is necessary for the successful negotiation of most tasks. The amount of such ability can be determined in a fairly accurate way for each vocation. Individuals who do not in this respect measure up to the minimum requirement in an occupation will do well not to attempt to enter it. The same is true with respect to higher education. The tests that have been given thus far seem to show that unless a person has the necessary fundamental basis it is undesirable for him to attempt to secure a higher education. At present we are applying crude methods of selection to nearly all of these activities. Oftentimes it may be merely vague personal opinions or chance peculiarities which form the basis of the judgment of the "expert."

---

[1] This paragraph was written in the fall of 1919. The prediction appears to have been justified.

The other phase of this problem concerns the specific abilities which are needed in special lines of work. A few tests have been devised which attempt to pick out the mental peculiarities of people who are successful in music, art, or other specific lines. We are making a beginning in this field and probably will make rapid progress from now on. Enough has been done in industry to indicate also that different occupations make their special demands. These specific requirements can be determined and individuals, who are not equipped with the peculiar capacities needed, can be rejected by the employment office. Thus, for example, one occupation may demand clear vision, another quick perception, and a third delicate motor adjustments. The degree to which each of these traits must be present to avoid probable failure may be established, and individuals not meeting the requirements for the occupation in question may be diverted from it.

The problems of vocational and educational guidance are much more complex than the problems of the employment manager. The expert in vocational or educational guidance is expected to make a wise recommendation for every individual who comes up for an analysis. The employment manager, on the other hand, usually has a number of people from whom he is privileged to select the best. This makes it possible for the tests used by the employment manager to have an element of error in them that would be fatal to the success of the test used by the counselor of individuals. The latter is most concerned with the future possibilities of the individual. Will the boy or girl who is receiving advice develop with further education in a way that will make his or her adjustment to the required conditions easy? The future must be considered to the extent of five or ten years. On the other hand, the employment manager is concerned with the immediate present. Only rarely will he consider the possibilities which may be attained by the individual five or ten years hence.

B—WHAT MENTAL TESTS MEASURE

I. *Phases of Mentality which are Measured*—In the popular mind there is much confusion as to just what mental tests measure. In general the thing sought to be measured by present mental tests is potential adaptability to conditions. How readily can the subject adjust himself to new situations? How quickly can he learn? To what extent can he profit by experience? It is a question of potential ability whether it arises from inherited native capacity or not.

From a more literal point of view, mental tests may be said to measure only the individual's performance. With his performance as a hypothesis we infer his ability. How near we come to the truth will depend upon how closely what he does corresponds to what he can do. In some cases the inference will not do the individual justice because he has not done his best—perhaps not nearly his best. But the standardization of procedure in giving tests and their repetition on different occasions with the same individuals will greatly reduce the likelihood of error in inferring ability from performance.

Moreover, we draw similar conclusions in regard to human behavior of all sorts. In other words, we infer ability from its outward manifestation in performance. A salesman's ability is gauged by the amount of his sales, a mechanic's ability by his visible product. The writer is judged by his books, the preacher by his sermons, the physician by his cures, and the business man by his holdings. In a world of action ability which does not eventuate in action is as if it were not.

Yet potential action—i.e. ability—must ever be in advance of actual performance. The margin between what can be done and what under given conditions is done varies between individuals and for the same individual at different times. When the conditions are favorable the margin is contracted and performance approaches the level of ability. Under unfavorable conditions performance may lag far behind ability.

How wide the habitual margin is for a particular individual is of little consequence. He may plead greater ability than he shows, but we shall continue to discount it to the level of his customary performance. Indeed, we may be theoretically as well as practically correct in so doing. It may very probably be true that a person, perhaps through emotional or volitional defect, exhibits a characteristic discrepancy between intellectual ability and action—a trait which is as peculiarly his own as his blue eyes or his aquiline nose, a trait in virtue of which an unusually large amount of his mentality cannot be brought into play. We are aware that in speaking of "amount of mentality" in this connection we are using a crude expression. It is only as the mentality permits action that we can speak of its "amount." Of what is over and above that which functions, we know nothing. It may be much or little, but since it accomplishes nothing further than to provide a working margin, we may safely neglect it in our tests of intelligence.

But the variation in this margin for the same individual at different times is serious. Unless we can allow for this variation or reduce it to a negligible amount, the reliability of our results is seriously impaired. A great deal of effort has, therefore, been expended in order that our results in terms of performance may become a usable index of mental ability. It is evident that the causes of unreliability—of the variability of this margin of which we have been speaking—lie in the changing conditions under which performance takes place.

These conditions are both external and internal. The external conditions include those of temperature, ventilation, illumination, and in general all the things which may at the time be present to the senses. In mental testing a set of especially important external factors has to do with the examiner. His directions may be clear or faulty, may give too much or too little information, may give a right or wrong "mental set." His voice may be entirely or but partly audible, harsh, or pleasing. His manner may be stimulating or depressing. Effort is made to reduce the variation due to these external conditions by standardizing them. This is especially true with regard to the examiner and the directions which he is to give. Some of the other external conditions—e.g. such variations in ventilation as are commonly found in schoolrooms—do not appear to make appreciable differences in performance. On the whole we believe that variations due to external conditions have been reasonably controlled where carefully devised tests have been properly used. Greater control is possible especially through more adequate training of examiners; and through the derivation of tests which require less special training on the part of the examiner. Progress is being made in both these directions.

Internal conditions under which performance takes place are only partly controlable. In the schoolroom we may, however, do more of this than is at first apparent. For example, the element of fatigue may be measurably controlled by a uniform schedule of work prior to the time of testing. On the other hand, remoter factors having to do with the condition in which children come to school are less easily controlled if indeed they can be controlled at all.

Unreliability due to variation in both external and internal conditions may also be reduced by repeating tests, by giving parallel tests, and by giving several different tests—in short by securing at different times additional data regarding the intelligence of the examinee. The extent to which this should be done in order to secure results of a given reliability is one of the promising statistical fields in which workers are now engaged but in which they have not, as yet, secured usable results. Meanwhile, however, it is evident that the reliability of a first determination of the mentality of an individual is greatly increased when no more than a single additional and independent determination is found to agree with it. Further determinations, if they are still substantially in agreement, will establish a degree of probability amounting to practical certainty. If determinations are not in reasonably close agreement, they may properly be regarded as chance variations from a presumably truer determination. The average of the ascertained determinations may be taken as the best representation of this truer determination.

In any event, therefore, inferring ability from performance is no new procedure. With care in administering mental tests, it is probable that we may make such inferences in reference to intelligence with reasonable accuracy.

Mental tests do not measure native capacity or general intelligence directly. They only indirectly get at these as they have been modified by experience. Even tests which are composed of the most perfect uncoachable elements are attempted more successfully by those who have had a thorough education than by those who have never been inside of a schoolroom. It is conceivable that tests may be devised which will measure pure intelligence—i.e. native capacity; but present tests are not of this nature, and it is questionable whether such tests are desirable. The individuals who are being measured are not the same individuals they would have been, if they had had different experiences. The important thing from all practical points of view is the present status of the individual. Theoretically, it may be interesting to compare two people on the basis of their pure native capacity, but for most practical purposes this question is unimportant. Experience plus native capacity has made the present individual, and it is he who must be considered.

II. *Phases of Mentality which are not Measured*—Mental tests do not measure the emotional side of life. They do not test one's ability to feel or to appreciate the finer things in art, nor do they test one's feeling of respect for one's fellows. They do not measure the ability to persevere, or to "carry on," except in a very limited way. Many people with mediocre endurance have sufficient power to enable them to work at a high pitch during the brief interval of a mental test, but they would be entirely unable to work twelve hours at a stretch, day after day. Mental tests do not measure the motives which guide the conduct of an individual—his conscience, his ideals, his honesty, and dependability.

This point of view, however, does not take into account the fact that these so-called emotional characteristics are apparently correlated in general with measurable mental characteristics. Usually the most brilliant individual from the mental point of view also has a very large endowment in ideals, endurance, persistence, and appreciation. In so far as the mental and emotional characteristics of human nature are correlated, tests of mental ability are also tests of emotional ability. The exceptions, however, are responsible for much of the criticism which is directed toward mental tests.

Again, mental tests do not measure directly the ability to use habits which have been acquired. Comparatively low-grade individuals may learn to do things which are mainly habitual activities. Through much practice they may have been perfected in the habits involved and, once having learned the habits, they may be able to practice them as effectively as the average individual. Consequently, mental tests so far as they test the performance of acquired habits may not be discriminative. Special tests are needed for this purpose. Mental tests, however, will indicate to a certain degree the speed with which individuals may acquire new habits and the facility with which they may modify old habits in new situations.

C—TYPES OF MENTAL TESTS

I. *Individual Scales.*

a. Characteristics of individual scales—The term "individual scales" is applied to those major measuring instruments which are used to test individuals one at a time. Such scales are sometimes called interview tests. They are composed of many items of very diverse character. These different bits of test material aim to determine the stage of development in the different functions involved in mental ability. The items differ in difficulty from very easy to very hard. The difficulty of each item is known with a reasonable degree of definiteness and the response to all the questions is combined in one value This value is commonly expressed in the form of a mental age.

In general, individual tests are regarded as our most accurate instruments. They have their limitations, however· If a pupil is sick or is unusually bashful, or becomes angry, the results are not descriptive of his real ability. Anything that prevents full cooperation with the examiner will invalidate the results. The different scales which have been devised for individual use have their own special limitations. A scale which is intended to measure children only between the ages of seven and fourteen should not be used in testing the ability of people whose mental age runs above these years. Further, a test which is merely a test of performance may not be a test of linguistic or other types of ability.

All the individual scales now in use require carefully trained examiners. Each item in the test will result in accurate information only after a careful following of directions and an accurate evaluation of the responses. A slight deviation from the standard wording of the directions will materially alter the response. Leniency or severity in the scoring of responses will influence the conclusions. Each answer must be evaluated accurately in the same manner that it was evaluated when the scale was devised. This can be done only by those who have made a thorough study of the scales and have some knowledge of child psychology. Some people are by temperament entirely unfitted to give individual examinations. Being unable to secure the cooperation of the subject, they obtain erroneous results· Thus an error of as much as two or three years may be made in the determination of the mental age of the subject.

The greatest objection, however, to the individual scales is that they are time-consuming. It takes from thirty minutes to two hours to administer either of the three scales described below to a single child. This fact makes individual methods of testing so expensive that they can never come into general use in the schools.

b. *Available individual scales*[2]—1. "The Point Scale" by Yerkes, Bridges, and Hardwick. The manual describing this scale is published by Warwick and York, Baltimore, Maryland. The scale is composed of 20 different tests, and the total possible score is 100 points. Within each test the items are graded somewhat in difficulty. Some of the tests are much easier than others but there is no careful gradation of the tests from very easy to very difficult. When this scale was devised it was intended to be valid between the mental ages of seven and fourteen. It has been found, however, in practice that if the scale is used with adults the results are questionable when the mental age exceeds twelve. The Point Scale may be given in less time than the two following scales, but it is a comparatively inflexible instrument. It deals primarily with literary material and places the unschooled individual at a decided disadvantage. The technique of its administration is somewhat difficult and no one should attempt to use it unless he has made a careful study of the manual and has been supervised in administering it. In other words, it cannot be given by an untrained examiner. The Point Scale is a modification of the early Binet-Simon Scale with the addition of a few new elements.

2. "The Stanford Revision of the Binet-Simon Scale" by Lewis M. Terman. The manual describing this scale is entitled "The Measurement of Intelligence" by Lewis M. Terman, published by the Houghton Mifflin Co. The envelope of test materials needed in administering the scale is furnished by the same publishers. This scale, as its name implies, is a revision and extension of the Binet-Simon Scale. It is composed of 90 different tests arranged in 12 groups corresponding to mental levels of from three to eighteen years. The large number of tests permits at least six of them to be included in each age group, thus securing a comparatively high reliability. The length of the scale, however, increases the time needed for its administration. Few workers who do thorough testing of children take less

---

[2] The materials needed for the administration of the three individual scales described are sold by C. H. Stoelting Co., 3047 Carroll Avenue, Chicago, Ill.

than an hour for each child when using this scale. In some cases it is necessary to use an hour and a half or two hours to complete the test according to instructions.

The scale is accompanied by very carefully prepared directions. Since it is composed of so many items, it can be administered successfully only by those who have had a thorough training. In comparison with the Point Scale it requires two or three times as much effort to learn to give the Stanford Revision. The results secured, however, are usually considered to be more significant. Individuals who have not gone to school, however, are penalized by the literary character of many of the tests and do not do themselves justice.

3. "A Scale of Performance Tests" by Rudolf Pintner and Donald G. Paterson, published by D. Appleton and Co., New York. The performance scale as devised by these authors has proved to be useful for measuring the mental ability of illiterates and foreigners. In the Psychological Service of the United States Army a modification of this scale was used with those men who could not be tested with the Point Scale or the Stanford Revision. As presented in this book, the scale is somewhat poorly adapted for school use. Modifications can be made, however, which will make it helpful in those situations where literary material cannot be used for test purposes.

II. *Group Scales.*

a. Characteristics—There are already a number of group scales for measuring the mental ability of children and adults. These are made up of several graded tests each of which is composed of individual items which are comparatively homogeneous. The theory underlying these tests is that several tests measuring different mental functions will measure general intelligence when the results are pooled. As at present arranged, it is felt that these group scales are not as accurate in their measurements as individual scales. In all probability, however, as high a degree of accuracy can be secured from the use of a number of group scales as from a single individual scale. This point has not been definitely settled at this time, however, and additional evidence is needed to guide us properly. Where a comparatively rough estimate of the mental ability of people is desired, these group scales answer the purpose very well. They have the decided advantage over the individual scales that they do not require much time per subject for their administration.

Moreover, most of them are not as complicated as the individual scales, and they can, therefore, be administered by intelligent people who have had comparatively little training. The points to be remembered in the administration of any one of the group intelligence scales are seldom as numerous as the instructions for two or three single tests in the Stanford Revision. It is very probable that group scales will supplant individual scales for general purposes. They will indicate those individuals who deviate from the norm and then, if the results need confirmation, it will be possible to give individual tests or to make additional studies of these unusual individuals.

b. *Existing group scales*—Nine important group intelligence scales have lately come to our attention. Doubtless there are others; for during the past six months a number of psychologists have been busy developing group scales. Under the auspices of the Bureau of Educational Research, six of these nine scales have been tried out during the past year. Detailed results will be given in Part II of this report. The six scales are the following:

1. "Otis Group Intelligence Scale," devised by Dr. Arthur S. Otis; published by the World Book Company.

2. "Classification Test," devised by Dr. W. W. Theisen and Mrs. Cecile White Flemming. Announced for publication by Teachers College, Columbia University.

3. "Group Test for Grammar Grades," devised by Professor Guy M. Whipple; published by the Public School Publishing Co., Bloomington, Illinois.

4. "Primer Scale," devised and published by Mrs. Luella W. Pressey, Indiana University, Bloomington, Indiana.

5. "Virginia Delta I," devised by Professor M. E. Haggerty for the Virginia Educational Commission; published by the World Book Company under the name *Intelligence Examination, Delta 2.*

6. "Sentence Vocabulary Scale," devised by the writer; published by the Bureau of Educational Research, University of Illinois, Urbana, Illinois.

The remaining three of the nine group scales to which we have referred were not tried out. They are briefly described

---

[3] Spring of 1919.

below. The first had been used earlier in the year at Danville and will be reported elsewhere. The second was for more advanced pupils than we were testing, and the third became available too late for use.

1. "Indiana Group Point Scale," devised and published by Sidney L. Pressey, Indiana University, Bloomington, Indiana. This group scale was one of the first to be published. As a pioneer scale it deserves no small credit, but it contains defects in administration which will prevent it from becoming popular in its present form. It is long and very exacting on the examiner; and the scoring is somewhat difficult. It cannot be given by teachers with success unless they have been carefully trained in its administration. The units of the scale are somewhat coarse and its discrimination is not very accurate. There are ten tests, each containing 20 items which are supposed to measure ability from the third grade through the high school.

2. "Psychological Examination for College Freshmen and High School Seniors, Parts A and B," devised and published by L. L. Thurstone, Carnegie Institute of Technology, Pittsburgh, Pennsylvania. This group scale is arranged in what is known as the "Omnibus Form." Its administration is exceedingly simple since the examiner has almost nothing to do except to start and stop those taking the test. The blanks contain complete directions. Little can be said by the writer as to the value of this scale. The materials used are approximately the same as those used in the Alpha Army Test. There is little doubt but that they are difficult enough for the groups of students (college freshmen and high-school seniors) for whom they have been devised.

3. "Virginia Delta VII for Grades I to III," devised by Professor M. E. Haggerty for the Virginia Education Commission; published by the World Book Company under the name *Intelligence Examination, Delta I.* This scale is one of the latest that has come to the attention of the writer. It seems to offer possibilities which will make it valuable for the primary grades. Nothing further can be said about it at this time since no published data are available.[4]

---

[4] Since this was written, standards have been provided. They are furnished when the tests are purchased.

PART II—COMPARISON OF GROUP MENTAL SCALES

A—INTRODUCTION

The rapid development of group scales has always been referred to. Some were planned before the United States entered the World War, and have been gradually developed since then. Others originated in connection with the work of the Psychological Service of the United States Army. Thus a number of group scales have become available for school use without much knowledge of their appropriateness for such use. Instead of considering each of these instruments from an *a priori* point of view, we have preferred to administer them to public school children under school conditions and to draw conclusions from the facts as thus revealed.

The opportunity to do this presented itself in connection with the work of the Bureau of Educational Research during the second semester of 1918-1919. With the cooperation of teachers and supervisors, six scales were administered to the school children of Champaign, Illinois, in the elementary and high school. The following scales were used: (1) Otis Group Intelligence Scale; (2) Classification Test, Form A; (3) Whipple's Group Tests for Grammar Grades; (4) Pressey Primer Scale; (5) General Examination No. 1—Virginia Delta I, and (6) Sentence Vocabulary Scales.

B—ADMINISTRATION OF TESTS

Approximately twenty-five hundred children were tested with one or more scales. With the exception of the Sentence Vocabulary Scales, which were administered by the individual teachers, all of the scales were given by the writer or by supervisory teachers or trained workers. Consequently, all of the data, with this single exception, may be considered to have been secured by disinterested people who could be relied upon to administer the scales according to instructions.

I. *The Otis Scale*—Due to physical limitations the administration of the Otis Scale was restricted to those grades in which it was thought that it could be given with greatest success. Furthermore, it seemed best not to attempt to give it to all of the schools in the city. Consequently it was offered only to grades VI to XII inclusive. In grades VI to VII the children in one average school alone were examined. In the eighth grade and in the high school the children were selected at random and were probably representative of these grades for the city. The first administration of the Otis Scale was in the eighth grade. The reaction of the children seemed to indicate that the scale would not be suitable for the lower intermediate grades. However, it was decided actually to give the test in the sixth and seventh grades to learn positively whether or not it was suitable for these grades. A general idea of the results obtained from administering the Otis Scale may be gathered from the central tendencies and variabilities for each grade as shown in Table I. The maximum score for this scale is 230.

TABLE I. TOTAL SCORES IN THE OTIS GROUP INTELLIGENCE SCALE

	Grade						
	VI	VII	VIII	IX	X	XI	XII
No. of pupils	47	17	61	99	95	93	54
Average	98	113	128	134	138	146	149
Standard deviation	15.5	23.5	20	25.5	25.5	24	22
Median	94	117	129	132	137	144	150

II. *Classification Test*—The number of pupils who could be tested by this scale was restricted by almost the same factors which limited the administration of the preceding scale. It seemed best not to test the same children as were tested with the Otis Scale. From the standpoint of the statistical study of the scales, it would have been desirable to measure identical children with them, but from the standpoint of the school system, it was felt to be better to test different children with each scale in order that a wider survey might result. In every case, however, at least one of the other four scales was given to each group of children examined by the Otis and Classification scales. It appeared likely that the Classification Test could not be used to advantage below the fifth grade. Accordingly, it was administered in grades V to XII inclusive, with the results indicated in Table II. In Table II are also given data obtained by Dr. Theisen from several Wisconsin communities.

TABLE II. TOTAL SCORES IN THE THEISEN-FLEMMING CLASSIFICATION TEST

	GRADES							
	V	VI	VII	VIII	IX	X	XI	XII
Champaign, Ill.								
No. of pupils	81	63	62	55	61	24	33	31
Average	60	84	98	108	118	125	134	141
Standard deviation	18.5	16	18.5	16.5	22	27.5	22	26
Median	59	83	99	110	116	131	136	141
Wisconsin [a]								
No. of pupils			142	101	608	289	118	262
Average			75	90	108	115	112	123
Standard deviation			22	21	24.5	29.5	20.5	24.5
Median			74	89	109	114	112	122

[a] Unpublished material furnished through the courtesy of Dr. W. W. Theisen.

III. *Whipple's Group Test*—It did not take very much work with this scale to show its general administrative inferiority in its present form to the other scales which were studied. It was found that much time was needed to give it and that a number of things about the scoring make it unsatisfactory from that point of view. The edition of the scale used at Champaign was the first offered by Dr. Whipple. It, therefore, contained defects which have been eliminated in later editions. For example, Test 5 was not printed in a form that was intelligible to the children, and it could not be used in this study. Moreover, there was no authorized procedure by which the scores in the different tests could be converted into a total score comparable to the total scores of other scales. It was administered to but 145 children. They were in grades IV to VI inclusive.

IV. *Pressey's Primer Scale*—The Primer Scale was administered throughout the city in grades I to III inclusive. The distribution of the scores is shown in Table III.

TABLE III. DISTRIBUTION OF TOTAL SCORES IN PRESSEY'S PRIMER SCALE BY GRADES

Score	Grade		
	I	II	III
0—4	1		
5—9	2		
10—14	2		
15—19	9	4	
20—24	8	1	
25—29	8	3	1
30—34	9	3	2
35—39	16	7	5
40—44	33	11	8
45—49	18	22	8
50—54	32	29	17
55—59	13	22	21
60—64	13	31	32
65—69	6	33	31
70—74		8	27
75—79		7	23
80—84		2	10
85—89			3
90—94			1
Total . . .	170	183	189
Average . . .	43.3	56.4	64.1
Standard deviation	14	12.5	12
Median . . .	44	57	64

V. *Virginia Delta I*—The materials for this test were furnished by Dr. M. E. Haggerty, Director of the Division of Tests and Measurements, Virginia Education Commission. That norms with which to compare the results of this test, which was being given in Virginia, might be available 1,200 copies were supplied to the Bureau of Educational Research. These proved sufficient to test all the children of Champaign, Illinois, in grades III to VIII inclusive. The results are shown in Table IV.

TABLE IV. TOTAL SCORES IN VIRGINA DELTA I.

Score	Grade					
	III	IV	V	VI	VII	VIII
0— 9	1					
10— 19	3					
20— 29	11					
30— 39	24	2				
40— 49	27	22				
50— 59	27	30	19	1		
60— 69	12	45	30	8	3	1
70— 79	7	47	51	20	6	0
80— 89	1	23	36	25	11	4
90— 99	2	9	34	32	32	16
100—109		7	17	31	29	33
110—119		1	8	28	40	38
120—129		1	4	10	23	33
130—139			0	3	19	38
140—149			2	1	3	15
150—159					1	1
160—169						1
Total	115	187	201	159	167	180
Average . . .	48	69.1	82.7	102.7	111.9	117
Standard deviation	16	16	16.5	16.5	17.5	16.5
Median . . .	48	68	80	98	110	117

VI. *Sentence Vocabulary Scales.*

a. Origin—The high degree of reliability of the vocabulary test contained in the Stanford Revision of the Binet-Simon Scale suggested that this material, if arranged as a group test, might prove valuable. Accordingly, sentences were devised each of which contained one of the words from the vocabulary lists of the Stanford Revision. The last word of each sentence was one of four words placed some distance to the right of the body of the sentence. The pupil taking the test was directed to underline one of the four words in each line which completed the sentence satisfactorily. The sentences were divided into two groups of fifty each on the same basis that Dr. Terman used in dividing his list. These two groups were called Series G and Series H. When thus arranged the sentences were mimeographed and were administered to several classes of children without a time limit. This preliminary use of the material revealed merit, and it was revised to remove obvious crudities of construction. The two series were then given to all the children in grades III to XII inclusive of the Champaign public schools. The papers were scored by deducting from the number of correctly underscored words one-third the number underscored incorrectly. This was done to reduce the effect of chance. Where a child underscored more than one word in a line, the sentence was counted as omitted.

b. Results—The analysis of the results soon revealed a wide deviation for individual pupils between the scores made in Series G and in Series H, although the median and average scores for the two series were about the same in a given grade. These differences were large enough to reduce the correlations between individual scores in the two series to surprisingly low values. In no grade was the correlation over +0.58. (See Table XXVI.)

When the results of two tests of the same kind show as much deviation for the individuals as this, the obvious thing to do is to combine the two scores into single indices. This was done, and the distribution of the total scores is presented in Table V. The deviations found between grade scores in Series G and in Series H led to a rearrangement of the two series. The total number of times that each of the one hundred sentences was completed correctly was computed for each grade. These results were reduced to percents, using as a base the total number of children who took the test in the grade in question. This procedure counts errors and omissions the same. With the resulting percents as a basis, the sentences were rearranged and are now presented as Series I and Series II. It is probable that these lists are as nearly of equal difficulty for the different grades as statistical computations can insure.

TABLE V. DISTRIBUTION OF TOTAL SCORES IN THE SENTENCE VOCABULARY SCALE

Score	Total Score for Grade									
	III	IV	V	VI	VII	VIII	IX	X	XI	XII
0—4	17	2								
5—9	18	1								
10—14	19	4	4							
15—19	27	10	11	2	1					
20—24	20	21	20	3	2					
25—29	15	33	32	19	2	1	1			
30—34	6	46	40	37	16	5	0			
35—39	7	21	39	28	24	11	6		3	
40—44	1	12	21	56	31	30	13	4	2	1
45—49		5	13	33	30	21	19	8	6	4
50—54		0	7	18	24	24	20	16	8	5
55—59		1	2	9	19	32	23	24	20	15
60—64		1		4	13	19	18	21	22	15
65—69				2	5	16	17	20	23	24
70—74					3	5	13	14	17	12
75—79						3	5	3	10	7
80—84							1	5	3	7
85—89								2	3	3
90—94								2	1	0
Total	130	157	189	211	168	167	136	119	118	93
Average	17.3	29	33.5	41.5	47.1	52.1	57.2	60.7	64.2	66.4
Standard-deviation	10.5	9.0	9.5	9.5	10.5	10.5	10.5	10.5	10.5	9.5
Median	21	31	33	40	47	53	56	61	64	66

Table VI shows the percent of pupils in each grade who responded correctly to each sentence. The percents are larger in most cases than they would be, if a deduction had been made for the number of times a sentence was underlined correctly by pure chance. Such a deduction was not made. If the chance factor were not present to inflate the percents, there would be an appreciable percent of pupils having zero scores. If every child had tried each sentence, the lowest percent theoretically would have been 25. It might be said in the light of the percents in this table that "a little knowledge is a dangerous thing" because the more mature high-school pupils, who tried to get some of the words by comparison of form and derivative roots, made more errors than the grade pupils who underlined purely at random.

One of the chief merits of the sentence vocabulary scale is the ease with which duplicate forms can be devised. The original 100 words contained in the Terman Vocabulary List were chosen by a random sampling method from the 1904 edition of Laird and Lee's *Vest Pocket Dictionary.* Other lists of 100 words can be selected by choosing words equally distant in the dictionary from those selected by Terman. The writer has already chosen the first and second words preceding Terman's and it is planned to present these in sentences at the first opportunity. Care should be taken in the derivation of duplicate forms to select the words in the sentences in such a manner that the ideas represented by the four completing words are approximately of the same degree of abstractness as the key-word in the sentence.

TABLE VI. PERCENT OF PUPILS MARKING VOCABULARY SENTENCES CORRECTLY

SERIES I

No. of Sentence	Grade									
	III	IV	V	VI	VII	VIII	IX	X	XI	XII
1	85	93	97	96	98	95	97	100	98	100
2	81	98	98	99	100	93	99	97	99	100
3	80	93	95	96	100	96	98	95	98	99
4	79	82	85	90	87	86	90	94	88	85
5	72	91	92	97	97	96	99	100	100	100
6	68	92	88	91	96	93	98	94	94	99
7	62	92	92	95	95	99	100	100	100	100
8	61	92	88	96	100	94	99	99	100	100
9	62	68	84	83	87	90	96	96	95	97
10	53	88	90	97	95	99	100	100	100	100
11	42	75	85	95	97	100	99	100	98	100
12	49	74	88	95	100	97	99	100	99	100
13	44	74	84	89	84	83	90	95	95	97
14	53	73	57	74	76	90	93	82	86	89
15	51	64	69	91	95	98	96	99	99	100
16	32	72	64	90	86	96	99	97	99	99
17	35	68	56	78	84	90	90	84	94	97
18	38	67	75	91	86	91	93	97	96	98
19	36	63	72	76	82	84	80	68	84	85
20	32	46	62	77	82	83	89	79	71	79
21	18	22	48	59	73	84	90	90	95	97
22	21	29	39	57	70	71	81	79	76	82
23	36	56	45	45	58	72	98	73	79	88
24	19	42	45	61	76	76	91	92	98	99
25	26	47	33	42	69	67	72	71	79	90
26	32	22	37	38	56	64	77	82	85	88
27	46	48	48	51	54	58	62	39	45	61
28	19	11	22	31	42	63	69	77	85	92
29	36	29	34	40	32	60	59	52	44	51
30	16	15	28	24	43	53	74	73	77	81
31	16	19	24	33	42	52	61	66	65	83
32	29	39	19	32	45	43	61	75	72	88
33	23	16	31	41	48	45	63	57	50	63
34	10	17	17	20	24	31	55	64	54	69
35	19	14	23	17	24	27	51	52	57	77
36	35	26	42	37	44	44	51	28	24	49
37	16	13	7	19	37	34	46	24	33	51
38	18	31	23	28	29	36	47	44	50	56
39	25	33	23	26	37	28	39	44	43	67
40	15	29	31	33	35	35	38	24	21	45
41	13	11	16	21	29	27	39	18	26	40
42	12	12	29	28	18	19	38	19	17	29
43	23	23	18	28	21	14	38	17	9	27
44	12	9	13	7	11	15	16	12	22	33
45	28	28	19	35	22	25	22	26	28	30
46	6	15	19	22	30	33	29	18	36	36
47	9	8	17	15	18	27	31	16	5	28
48	30	37	38	23	22	30	36	19	5	27
49	7	23	20	14	11	17	4	13	10	15
50	14	14	17	7	20	15	34	10	10	14

TABLE VI—(*Continued*)

SERIES II

No. of Sentence	Grade									
	III	IV	V	VI	VII	VIII	IX	X	XI	XII
1	96	98	100	99	99	89	100	100	100	100
2	81	98	96	98	95	99	100	100	100	100
3	84	89	90	94	96	100	99	98	100	100
4	77	97	93	99	100	98	99	100	99	99
5	72	95	95	97	100	97	99	100	99	100
6	68	91	86	96	100	94	99	99	98	100
7	63	79	79	89	93	92	93	94	94	99
8	60	87	86	97	96	93	97	91	98	95
9	58	82	84	95	95	97	99	98	100	100
10	58	94	88	95	97	98	99	100	98	99
11	51	77	78	82	90	94	96	98	95	96
12	59	78	69	88	95	91	97	91	99	98
13	55	72	69	87	93	94	96	91	96	98
14	48	72	81	78	88	94	99	99	100	100
15	49	72	83	91	96	93	98	100	97	95
16	35	68	75	85	90	94	99	100	98	100
17	40	65	75	79	87	93	98	94	93	97
18	41	67	67	80	77	85	92	84	90	87
19	22	58	46	74	87	80	86	83	95	95
20	27	48	66	67	82	88	86	98	95	100
21	13	35	32	63	68	66	89	71	76	73
22	35	26	40	49	67	84	80	70	79	89
23	19	41	40	57	79	84	96	96	97	98
24	12	32	32	53	67	60	82	86	88	89
25	7	12	26	34	62	79	91	85	92	92
26	33	37	39	40	59	66	88	87	88	85
27	42	41	48	42	54	49	55	46	50	55
28	6	7	13	15	45	76	79	91	98	99
29	7	17	33	26	36	55	71	66	67	86
30	14	15	26	26	47	54	64	72	75	84
31	6	19	15	31	37	52	61	76	78	81
32	16	5	13	14	44	43	61	63	61	85
33	14	26	24	24	37	42	64	57	61	63
34	18	47	26	47	48	48	57	70	85	87
35	19	14	13	13	32	24	54	40	39	54
36	16	22	20	19	37	29	53	48	38	48
37	19	19	15	25	36	24	44	30	33	53
38	23	11	16	17	26	33	46	53	61	71
39	24	32	23	19	23	27	29	30	43	47
40	20	27	25	27	41	35	51	22	22	41
41	12	5	6	4	9	8	26	32	42	63
42	32	28	35	25	34	14	40	16	19	19
43	19	25	19	17	15	20	19	29	21	35
44	14	28	20	28	24	32	32	18	24	35
45	32	26	21	29	36	20	19	16	10	30
46	10	13	10	3	46	11	25	12	16	31
47	15	33	34	43	32	35	34	23	20	29
48	11	23	22	16	21	18	22	16	5	22
49	15	32	15	22	22	17	23	22	2	17
50	12	7	15	5	4	8	5	6	10	14

C—THE TIME ELEMENT

One of the important considerations in selecting scales is the time needed for their administration and evaluation. Some group scales are so arranged that it takes a minimum of time to give them as well as to score them; while others go to the other extreme. Table VII presents briefly the approximate amount of time required to give those considered in this study.

TABLE VII. TIME CONSUMED IN ADMINISTERING THE SCALES

SCALE	MINUTES
Otis	70
Classification . . .	50
Virginia Delta I . .	30
Primer	25
Whipple's Group . .	80
Sentence Vocabulary	20—40[a]

[a] This scale is given without a time limit; but the time varies from about forty minutes in the third grade to twenty minutes or less in the twelfth.

These times are approximately those used in administering the scales at Champaign. Of course, much time can be wasted by inefficient routine methods. The papers should be distributed and collected quickly. No petty interruptions should be permitted while the pupils are working. Under these conditions the time required for the administration of the different scales will be approximately as indicated above.

After the intelligence scales have been administered much time is needed for scoring the papers. Many do not realize the tediousness of this work. It often costs more to score the papers and evaluate the results than to purchase the test materials. Data are presented here to show the approximate rate at which the scales used in this study were scored. This work was done by trained clerks who used stencils wherever possible. Few teachers will approximate this rate of work when they first attempt to score similar papers. The numbers of papers scored per hour by our clerks are indicated in Table VIII for the different scales.

TABLE VIII. RATE OF SCORING BY CLERICAL WORKERS

SCALE	NO. SCORED PER HOUR
Otis	13
Classification	15
Virginia Delta I	20
Primer	35
Whipple's Group Test . . .	6
Sentence Vocabulary (one series)	40

The time which may be devoted to the analysis of data after papers are scored is exceedingly variable. It depends to a large extent upon the purpose to be served. However, approximately the same amount of time should be allowed for this as is needed for scoring. With the exception of Whipple's Group Test all the scales yield a total score. This fact places them on an equal basis after the total scores have been obtained. In other words, from this point on, each scale will require about the same amount of time—unless an analysis is made of some or all of the individual tests which compose the scales. This last process will generally be unimportant for school purposes.

D—COMPARISONS OF TOTAL SCORES

I. *Correlations with Scholarship*—The correlations between the intelligence scores and scholarship are shown in Table IX. These values at first glance would seem to imply that the scales are not very reliable, that they do not adequately measure the mental characteristics important for school success. This might be the case, if the judgments of scholarship were entirely adequate. It will be worth while to consider this point briefly.

The teachers were instructed to rate the children in scholarship on a special sheet. Accompanying this sheet was a set of mimeographed instructions which directed that letter ratings should be so distributed that the teacher of a normal class would give 5 percent of the class A's, 20 percent B's, 50 percent C's, 20 percent D's, and 5 percent E's. If a group was abnormal, the teacher was asked to rate the children in comparison with all children of the same sex, race, and age. If these instructions had been carefully followed, the correlation with each scale would have been higher and more significant.

TABLE IX. CORRELATION BETWEEN SCORES IN THE INTELLIGENCE SCALES AND TEACHERS' SCHOLARSHIP RATINGS

GRADE	SCALE				
	Primer	Vocabulary	Va. Delta I	Classification	Otis
I	.42±.04				
II	.30±.05				
III	.30±.05	.54±.05	.57±.06		
IV		.50±.04	.45±.05		
V		.42±.03	.56±.04	.67±.03	
VI		.50±.04	.69±.03	.56±.06	.33±.07
VII		.45±.04	.71±.04	.51±.06	[a]
VIII		.46±.04	.58±.05	.28±.07	.41±.07
IX		.42±.05		.57±.06	.49±.05
X		.27±.06		.23±.13	.36±.06
XI		.59±.04		.56±.10	.55±.05
XII		.53±.04		.64±.10	.41±.07

[a] Cases were too few to be significant.

The teachers, however, were not able to follow the instructions very closely. This fact is shown by Tables X and XI which present the distributions of the scholarship rating for the first and fourth grades respectively, these grades having been taken as typical. Among the different schools it is evident even without converting the number of ratings into percents that there are wide deviations from the suggested percentage distribution of rating. Even when the ratings for all the schools are combined and converted into percents the discrepancy between the actual and theoretical distribution is still evident. It is clear, for example, that the first- and fourth-grade teachers gave a great many more A's than would have been expected. The number of C's was appreciably below the standard number, while the numbers of B's and D's (at least in the first grade) were of about the right order of magnitude.

The average grade for each school was computed by allowing the customary ratings of 5, 4, 3, 2, and 1 respectively for the letters A, B, C, D, and E. These averages conceal a great deal. A teacher for example, may give too many A's but if she balances them by giving too many E's the average may turn out to be 3 and the impression may be created that the distribution was correct. Nevertheless the averages do serve to indicate whether there is a constant bias on the part of the teacher in question in virtue of which she rates everybody too high or too low. There is, in the first grade, a slight tendency for teachers to rate their children above 3—that is above the expected average. This tendency is considerably more marked in the fourth grade. These deviations from the expected average are not at all accounted for by the scores in the mental tests at the schools in question. Median scores in the Primer Scale are shown in Table X and those for the Vocabulary Scale are shown in Table XI.

TABLE X. DISTRIBUTION OF TEACHERS' SCHOLARSHIP RATINGS FIRST GRADE

RATING	NUMBER OF RATINGS AT SCHOOL:						RATINGS AT ALL SCHOOLS	
	1	2	3	4	5	6	No.	Percent
A	2	6	8	6	4	7	33	17
B	3	14	5	0	7	7	36	19
C	17	7	8	14	11	7	64	33
D	8	7	8	8	6	1	38	20
E	3	4	4	3	4	3	21	11
Average	2.8	3.3	3.1	3	3	3.5	3.1	
Median Score Primer Scale	50	48	40	43	41	46	44	

TABLE XI. DISTRIBUTION OF TEACHERS' SCHOLARSHIP RATINGS FOURTH GRADE

RATING	NUMBER OF RATINGS AT SCHOOL:						RATINGS AT ALL SCHOOLS	
	1	2	3	4	5	6	No.	Percent
A	10	6	13	4	0	1	34	18
B	3	8	16	17	5	4	53	29
C	19	14	8	10	13	5	69	37
D	5	7	3	2	3	2	22	12
E	0	1	0	0	6	1	8	4
Average	3.5	3.3	3.9	3.7	2.6	3	3.4	
Median Score Vocab. Scale	29	33	30	32	28	30	32	

The effect of these individual variations on the scholarship ratings by the teachers is to lower the coefficients of correlation. But another factor which diminishes the correlations has to do with effort. It is well known that many school children are working far below the limit of their ability. If a superior child does not apply himself his scholarship rating may be mediocre or even poor, although his intelligence score may be high. Indeed, the novelty of the test situation and the shortness of the effort required frequently combine to secure a performance which corresponds more closely to actual ability than does the sustained routine performance of the classroom. Again, the exceptional industry of a child of ordinary ability may place him among the best, or at least much above his companions of equal general ability. Other factors, such as sickness, irregular attendance, and change of schools may cause a child's scholarship to be rated much below what it would be under normal conditions.

Under these circumstances, the distribution of the coefficients of correlation for the different tests reveal a reasonably high correspondence with scholarship ratings. (See Table XII.) The average for the entire group is +0.462. If the correlations for the Primer Scale are omitted, the average is a little higher, namely, +0.497.

TABLE XII. DISTRIBUTION OF COEFFICIENTS OF CORRELATION BETWEEN INTELLIGENCE SCORES AND SCHOLARSHIP RATINGS

COEFFICIENTS	NUMBER
.70—.79	1
.60—.69	4
.50—.59	12
.40—.49	9
.30—.39	4
.20—.29	3
Average	.462

It was thought that it might be significant to combine the scores of two somewhat dissimilar scales like the Virginia Delta I and the Sentence Vocabulary.[5] The combined scores of these two scales ought to show a higher correlation with scholarship then either alone, if they measure different phases of intelligence accurately, and to the extent that the ratings of scholarship are reliable. The resulting coefficients of correlation are as follows: Grade III, 0.64; Grade IV, 0.44; Grade V, 0.52; Grade VI, 0.54; Grade VII, 0.52; and Grade VIII, 0.54; Average 0.53. These figures do not show that the combination has very materially raised the correlations.

Our evidence as well as that exhibited in other investigations tends to show that while scholarship and intelligence are by no means independent, their relationship is also by no means perfect. Each is affected by conditions which do not affect the other to the same degree. Even if scholarship is accurately judged by teachers, perfect correlation cannot be expected. It is probable that a coefficient of much more than +0.60 between mental test results and estimates of scholarship would mean either that the test or the estimates were faulty. The test might be such that success in it depended too much upon schooling; or the estimates of scholarship might be too greatly influenced by the notion of natural ability.

---

[5] The scores were combined by dropping the score made in Test 3 of the Virginia Delta I, multiplying the Sentence Vocabulary score by two and finding the total. This procedure attempted to give equal weight to both of the scales. Test 3 was dropped because it is not discriminative.

II. *Correlations between the Scales*—The correlations between the different scales administered in this study are interesting and suggestive. As many as the data permit are shown in Table XIII. In some cases results from the same grade in several schools were used; in others the correlations had to be determined for the grade of one school only. Determining these values for a single class applies a much more rigid standard to a scale than would be the case, if correlations were computed from the combined results for different classes in the same grade or for different grades. Note, for example, that the correlations of the Sentence Vocabulary Scale with the Virginia Scale Delta I are without exception higher for all schools (column 2) than they are for one school (column 3). This is significant. In all probability it implies that the true correlation between the Sentence Vocabulary Scale and Virginia Delta I is appreciably higher than is here indicated.

In general, higher correlations are found where results from several different grades are used. This is because such a selection gives a greater spread of abilities. The two combination correlations presented in Table XIII for the Otis Scale reveal this tendency. (See the last entries in columns 5 and 7.) This procedure has sometimes been adopted in studies of mental tests. But a high correlation of this sort is not so significant as a high correlation secured from the more homogeneous material of a single grade. Since intelligence scales will be most useful, if they distinguish between the children of a single grade, this rigid test will be employed in the consideration of the merits of the different scales. This criterion should not be confused, however, with the procedure of others who compute correlations by using data from the combination of several grades.

TABLE XIII. CORRELATIONS BETWEEN SCALES [a]

GRADE	SENTENCE VOCABULARY WITH:					VIRGINIA DELTA I WITH:	
	Primer	Virginia Delta I		Classification	Otis	Classification	Otis
		All Schools	School 1				
	1	2	3	4	5	6	7
III	.40±.06	.89±.02	.86±.03				
IV		.60±.04	.53±.09				
V		.66±.05	.59±.09	.52±.06		.69±.04	
VI		.62±.02	.60±.09	.51±.05	.64±.06	.72±.04	.68±.06
VII		.78±.04	.73±.06	.33±.07	b	.81±.03	b
VIII		.69±.03		.76±.04	.67±.04	.65±.04	.89±.02
IX				.78±.04	.72±.03		
X				.37±.09	.53±.09		
XI				.67±.05	.77±.03		
XII				.62±.10	.62±.06		
Average		.71	.66	.57	.66	.72	.78
VI, VII, VIII					.79±.02		.82±.02

[a] No correlations are given for Otis with Classification nor for Virginia Delta I with Primer because in neither case did the same group of pupils take both tests.

[b] Too few pupils took these tests to warrant calculating the coefficients.

The coefficients presented in Table XIII are significantly high. If all of the correlations (except those of column 3 and those for combined grades) are distributed, they are as shown in Table XIV.

TABLE XIV. DISTRIBUTION OF INTER-SCALE CORRELATIONS

Correlations	All Tests
.80—.89	3
.70—.79	6
.60—.69	12
.50—.59	3
.40—.49	1
.30—.39	2

The high correlation values that have been obtained where two scales have been given to the children in a grade seem to indicate that these scales in the main are measuring much the same thing. When the diverse character of the measuring instruments is considered this fact is somewhat remarkable. Of course, the values are all low enough to lead one to be somewhat conservative in drawing conclusions concerning individual pupils from a single test.

III. *Reliability of Total Scores*—The correlation discussed in the two preceding sections may be considered to indicate the worth of mental tests from the point of view of relationships. These correlations, however, do not reveal many of the things which one would like to know about the different intelligence scales. It may be a valuable thing to have tests which correlate highly with one another or it may be the reverse. Further, it may be worth while to have tests which correspond closely to the scholarship ratings made by teachers; but on the other hand, if the scholarship ratings, given by the teachers under the conditions described above, do not forecast the real possibilities of pupils, high correlation with such scholarship ratings may not be either desirable or informing. It is conceivable that mental scales may be devised which to a greater extent than is true of these scales will direct attention to pupils who are brilliant, average, or mediocre, in a way that will enable teachers to develop their talents. It may be that the methods of education now in use are pedantic or that intelligence scales measure qualities which are highly desirable in life, but which do not function in school work. These questions cannot be settled here; but it will be worth while to consider the scales from another point of view, namely that of their power to discriminate between different intelligence levels.

Much of the discussion from this point on will make the assumption that there is a difference in intelligence level between grades—that third-grade children are, in general, more mature in intelligence than second-grade children, that fourth-grade children are still more mature than third-grade children, and so on. As the higher grades are reached, this difference probably decreases in absolute value. Correspondingly, the grade intervals shown by a test at different grade levels should probably not be equal. The differences, however, are taken to be appreciable; and good tests ought to reveal them clearly. Further, it should be possible to magnify these amounts somewhat by the use of scales especially devised to discriminate at the higher ranges. This assumption is approximately that of Binet, Terman, and others who developed the individual intelligence scales. There is this difference, however, the earlier workers made their groupings upon an age basis. The groupings here made are upon the school-grade basis. This basis was used, first, because the age groups are not completely represented in school. The least intelligent children of ages five, six, and seven have sometimes not entered school, while the more brilliant ones beyond the age of fifteen have in many instances, completed the public school· Consequently, we can use twelve unselected grade groups, while we would not have that number of comparable age groups. The grade basis was used, second, because it is more serviceable. In the schools children are classified by grades, not by ages. If a test is given, it is given to the children of the same grade, not to those of the same age. Standards for the grades and differences between grades are therefore more immediately useful than similar facts on the basis of age.

With the assumption made in the preceding paragraph we may compare the different intelligence scales on the basis of their power to discriminate between grades. If the fourth grade is more mature than the third there should be a difference between the third grade and the fourth grade in intelligence scores. That scale which reveals the most reliable differences between averages for the different grades may be considered to be the most discriminative. These facts have been computed for the different scales and are presented in Table XV.

TABLE XV. DIFFERENCES BETWEEN GRADE AVERAGES WITH RELIABILITIES (PROBABLE ERRORS) FOR TOTAL SCORES IN EACH SCALE

GRADES	DIFFERENCES BETWEEN AVERAGES						
	Primer	Vocabulary	Otis	CLASSIFICATION		Virginia	Virginia & Vocabulary
				Wisconsin	Champaign		
I & II	13.1±0.9						
II & III	7.7±0.9						
III & IV	……	11.7±0.8	……	……	……	21.1±1.2	39.7±3.3
IV & V	……	4.5±0.7	……	……	……	13.6±1.1	19.4±2.3
V & VI	……	8.0±0.6	……	……	23.7±1.9	20.0±1.2	28.5±2.3
VI & VII	……	5.6±0.7	15.2±4.1	……	13.5±2.1	9.2±1.2	24.2±2.6
VII & VIII	……	5.0±0.7	15.0±4.1	15.3±1.7	10.7±2.1	5.1±1.2	17.7±2.8
VIII & IX	……	5.1±0.8	5.5±2.5	17.5±1.6	10.1±2.4		
IX & X	……	3.5±0.9	4.7±2.5	7.3±1.3	6.2±3.6		
X & XI	……	3.5±0.9	7.3±2.5	−2.4±1.7	9.6±4.0		
XI & XII	……	2.2±0.9	3.3±2.6	10.2±1.6	7.3±4.0		

Since the different scales involve different numbers of units, a unit of one (e.g. the Vocabulary Scale with a possible score of 100) is not equivalent to a unit of another (e.g., the Otis Scale with a possible score of 230). The differences, however, between the grade averages shown in Table XV may be reduced to the same basis, either by dividing the grade differences by their probable errors (thus making the differences comparable), or reciprocally, by dividing the probable errors by the differences. This latter procedure gives us the so-called coefficient of variability. It is commonly held in statistical circles that a quantity should be at least three times its probable error to be worthy of statistical consideration, or, in other words that the probable error should be no more than one-third of the quantity.

With this standard we may examine the coefficients of variability presented in Table XVI remembering that the smaller they are the more reliable are the grade differences given in Table XV and that according to the standard they should be less than 0.333. Comparing the values grade by grade, it can be seen readily that some of the scales are much more discriminative than others. The best values for grades III to VI are shown by the Virginia Scale. From grades VI to XII the Vocabulary Scale seems to show the best discrimination. It is somewhat unfair to compare the scores secured in Champaign with the results furnished by Dr. Theisen for the Classification Test in Wisconsin, for it is altogether possible that a measurement of the children in Wisconsin with the Otis and Vocabulary scales would reveal a greater degree of discrimination than these tests revealed in Champaign. However, the comparatively good values shown by the Wisconsin figures are interesting and suggestive. To a certain extent the large (and hence unfavorable) coefficients yielded for the Otis and Classification scales by the Champaign data are due to the fact that the number of children tested was less than 100 in every grade. But this is not the sole reason. The numbers of children who took these tests were no larger in the grades below the high school, yet the coefficients of variability are much smaller for those grades. In the high-school grades only one of the eight coefficients of variability for the Otis and Classification scales in Champaign is less than 0.333. This would imply, so far as our data permit an inference, that neither test is as reliable for high-school work as one would wish.

TABLE XVI. COEFFICIENT OF VARIABILITY OF DIFFERENCES BETWEEN TOTAL SCORES FOR SUCCESSIVE GRADES

GRADES	PRIMER	VOCABULARY	OTIS	CLASSIFICATION		VIRGINIA	VIRGINIA & VOCABULARY
				Wisconsin	Champaign		
I & II	0.068						
II & III	0.117						
III & IV	....	0.068	....	....	....	0.056	0.083
IV & V	....	0.155	....	....	....	0.108	0.118
V & VI	....	0.075	....	....	0.080	0.060	0.080
VI & VII	....	0.125	0.269	....	0.155	0.130	0.107
VII & VIII	....	0.140	0.273	0.111	0.196	0.235	0.158
VIII & IX	....	0.156	0.454	0.091	0.237		
IX & X	....	0.257	0.531	0.178	0.580		
X & XI	....	0.257	0.342	0.708	0.416		
XI & XII	....	0.409	0.787	0.156	0.547		

The discriminative power revealed by the scales for wider ranges of grades is also interesting. A few of the many possible facts of this sort are presented in Table XVII. Observe that the grades set up in this table have reference to the elementary school, the junior high school, and the senior high school. The greatest discriminative ability is shown by the sum of the Virginia and Vocabulary scales. The Virginia Scale by itself seems to be twice as discriminating as the Vocabulary Scale for the elementary school. When the Vocabulary, Otis, and Classification scales are compared for grades VI to IX, they rank in that order. For the last three years of the high school the Classification Test is first, the Vocabulary second, and the Otis third. These facts imply not only that some of the scales are more discriminative than others, but also that we may secure better scales than any we now have.

E—ANALYSIS OF THE INDIVIDUAL TESTS IN THE INTELLIGENCE SCALES

I. *Individual Test Scores*—Each of the instruments used in this investigation for measuring intelligence (except the Sentence Vocabulary) consists of several series of questions or things to do. Each of these instruments without regard to its actual title may be called a *scale;* and each separately organized group of questions and things to do may be called a *test.* Thus, the Otis Group Intelligence Scale has ten tests; Whipple's Group Test for Grammar Grades has six tests; the Virginia Delta I has six tests; the Classification Test has eight tests; and the Primer Scale has four tests. Since the total score on a scale is to be taken as indicative of intelligence, the theory is that each of the component tests shall "tap" important elements of intelligence, and that the score on each test shall enter into the total for the scale to a degree that will give the proper emphasis to the element or phase of intelligence to which the test relates.

TABLE XVII. DISCRIMINATION BETWEEN LARGER GRADE RANGES FOR INTELLIGENCE SCALES

SCALE	GRADES III TO VIII		GRADES VI TO XI		GRADES IX TO XII	
	Difference between Averages	Coefficient of Variability	Difference between Averages	Coefficient of Variability	Difference between Averages	Coefficient of Variability
Vocabulary	34.8±1.2	0.034	15.7±0.7	0.044	9.2±0.9	0.098
Virginia	69.0±1.3	0.019				
Virginia and Vocabulary	129.5±0.3	0.003				
Otis			35.7±2.3	0.064	15.3±2.1	0.137
Classification:						
Champaign			34.3±2.3	0.067	23.1±2.1	0.091
Wisconsin					15.1±1.2	0.079

The tests of which each scale is composed are as follows:

Otis Group Intelligence Scale

I. Following Directions
II. Opposites
III. Disarranged Sentences
IV. Proverbs
V. Arithmetic
VI. Geometric Figures
VII. Analogies
VIII. Similarities
IX. Narrative Completion
X. Memory

Classification Test

I. Following Directions
II. Synonym-Antonym
III. Arithmetic
IV. Common Sense
V. Completion
VI. Analogies
VII. Number Completion
VIII. Information

Virginia Delta I

I. True-False
II. Arithmetic
III. Picture Completion
IV. Synonym-Antonym
V. Common Sense
VI. Information

Whipple's Group Test for Grammar Grades

I. Arithmetic
II. Completion
III. Substitution
IV. Reasoning
  Part I
  Part II
  Part III
V. Punched-Hole Test
VI. Proverbs

Pressey Primer Scale

I. Dot Pattern
II. Classification
III. Form Board
IV. Absurdities

Now, it will probably occur to anyone who has considered the foregoing material thoughtfully that the various tests which compose the different scales are not likely to be of the same value. To determine the extent to which this is the case, distributions were made for each test by grades. These distributions and interpretations were enlightening, but the limitations of this bulletin permit the presentation of no more than the general features.

A marked difference in the curves of distribution was shown by the several tests. Differences were also shown when the school-grade distributions were compared with each other for tests of the same kind occurring in different scales. In other words, some of the tests were not suited to the grades in which they were used, were too hard or too easy or too irregular. These facts will be presented more in detail in the next section.

II. *Differences betwen Successive Grade Averages*—The differences between the grade averages for each of the tests were computed. These revealed wide divergencies, showing that some of the tests were poorly adapted to the work that was expected of them.

The Otis Scale contains tests which are of small diagnostic value when discrimination between successive mental levels is sought. A number of inversions, cases where the score in a test decreased with the next higher grade, were noticed. While these might be due to some degree to an insufficiency of cases, this would not fully explain the fact that some tests showed inversions while others did not. The inversions suggest poorly constructed tests. For each test of the Otis Scale the differences between the average score in the highest and lowest grade was determined. It was thus found that the amounts contributed to the difference between the scores for the entire scale varied markedly, being seven times as large for Test IV (Proverbs) and Test IX (Narrative Completion) as for Test VIII (Similarities). The amount of this difference for each test—i.e., of the difference between the average score of the lowest and highest grades concerned—may be taken as one indication of the discriminative power of the test. The tests in the Otis Scale arranged in order from the most discriminative to the least according to the differences between the sixth and twelfth-grade averages are given in Table XVIII. As a whole, the Otis Scale showed more discriminative power in the grammar grades than in the high school.

TABLE XVIII. DISCRIMINATIVE POWER OF TESTS. OTIS SCALE

TEST	DIFFERENCE BETWEEN SIXTH AND TWELFTH GRADE AVERAGES	PROBABLE ERROR OF THIS DIFFERENCE
IV	8.5	0.5
IX	8.4	0.7
III	8.2	0.5
II	6.7	0.5
VII	4.3	0.5
VI	4.0	0.4
V	2.6	0.4
X	2.0	0.4
I	1.8	0.4
VIII	1.2	0.4

The Classification Test is better organized than the Otis Scale when the power of the individual tests to discriminate between successive grades is used as a criterion. Although no more of the Classification Test were administered than of the Otis, there were in the former case fewer negative differences between the averages of successive grades.

Moreover, judged by the larger recorded differences between the performance of sixth- and twelfth-grade children, the component tests of the Classification Test are considerably were found as shown in Table XIX. Again, taken as a whole, this scale, like the Otis, is more discriminative in the grammar grades than in the high school.

TABLE XIX. DISCRIMINATIVE POWER OF TESTS.
CLASSIFICATION TEST

Test.	Difference between Sixth and Twelfth Grade Averages	Probable Error of This Difference
VIII	12.2	0.6
II	11.9	1.0
VI	9.9	1.0
V	7.2	0.5
III	6.3	0.4
I	5.0	0.3
VII	3.8	0.4
IV	3.6	0.3

Marked differences between the discrimination of tests of the Virginia Delta I Scale are also evident. Indeed, the incomplete picture (Test III) is apparently not suitable above the primary grades. Arranged in the order of the differences between averages for grades III to VIII the tests are as follows: Test VI, 21.9±0.4; Test I, 14.8±0.5; Test IV, 14.7±0.4; Test V, 8.2±0.1; Test II, 6.3±0.1; and Test III, 1.1±0.2.

There is very little difference in discrimination between the tests of the Pressey Primer Scale or between the two vocabulary scales. Such differences as exist do not materially affect the use of either of these measuring instruments.

In general it may be said that the present scales have been arranged without a careful analysis of their component parts. They are poorly balanced when the individual tests are considered, and the comparative success that attends their use at present is due more to the homogeneity of human mentality than to the scientific derivation of the measuring instruments.

III. *The Coefficients of Variability of Individual Tests*—The influence of different numbers of units in the various individual tests may be eliminated by reducing the values, which have been the basis for the discussion in the two preceding sections, to coefficients of variability in the same way that we calculated the coefficients of variability for differences in total scores. There is one point which must be considered, however, and that is that the probable error of the difference between two averages is affected in a constant direction by the number of cases involved. Although the differences between any two averages probably will not vary markedly with double the number of cases, the probable error of that difference will be smaller, if twice the number of cases are studied. Where varying numbers of pupils have been tested by the scales, this fact introduces a certain element of unreliability into any comparisons which may be made between the different scales. This point applies to total scores as well as to the scores in the individual tests, although it was not mentioned in the section discussing total scores. Within the same scale, however, this fact ceases to operate because the same number of cases is involved in all the tests of a scale.

The coefficients of variability for the individual tests of the six scales studied are presented in Tables XX to XXV. The wide variations between the individual tests which were noted briefly in our discussion of the differences between successive grade averages have disappeared. The result of reducing them all to the same basis without reference to their effect on the total score shows them to be more nearly alike in discriminative power if the differences in weighting were equalized. Nevertheless, it can still be shown that some of the scales are much better organized than others. In the Otis Scale, for example, Test I shows a coefficient of variability for grades VI to IX (0.78) that is ten times as large as that shown by Test VI (0.076). An extreme range, such as this, is plainly not to the advantage of a scale. The variations in discriminative power between tests should not be considered as evidence that the tests are of no value. They merely mean that in the present form the tests are not well organized for the work that they are to do. It is altogether possible that a revision might introduce the changes that are needed.

TABLE XX—COEFFICIENTS OF VARIABILITY OF INDIVIDUAL TESTS BY GRADES FOR THE OTIS SCALE

Test	Coefficients of Variability of Differences between Scores for Indicated Grades								
	6&7	7&8	8&9	9&10	10&11	11&12	6&12	6&9	9&12
I	7.0	0.315	0.285	0.50	0.50	3.0	0.055	0.78	0.37
II	0.50	0.50	0.235	0.666	0.363	0.571	0.074	0.099	0.17
III	0.80	0.296	4.0	0.40	0.125	0.357	0.060	0.127	0.104
IV	0.379	0.478	1.666	0.166	2.50	3.0	0.058	0.078	0.19
V	0.277	5.0	0.187	0.50	0.375	0.363	0.153	0.096	0.397
VI	0.250	0.454	0.50	0.30	1.0	0.30	0.025	0.076	1.10
VII	1.20	0.137	0.60	0.750	1.50	2.0	0.116	0.086	0.854
VIII	5.0	5.0	0.272	0.375	0.333	1.50	0.333	0.322	3.37
IX	1.0	0.272	1.750	0.857	0.60	0.714	0.033	0.123	0.224
X	0.625	0.312	0.666	3.0	3.0	[a]	0.20	0.213	1.85

[a] The difference was zero. This made it impossible to compute a coefficient of variability that would mean anything to any but a mathematician.

TABLE XXI—COEFFICIENTS OF VARIABILITY OF INDIVIDUAL TESTS BY GRADES FOR CLASSIFICATION TEST

Test	Coefficients of Variability of Differences between Scores for Indicated Grades								
	5&6	6&7	7&8	8&9	9&10	10&11	11&12	5&12	6&9	9&12
I	2.0	0.350	0.692	1.0	0.60	0.428	0.50	0.056	0.062	0.016
II	0.108	1.40	0.080	0.380	0.75	0.333	4.33	0.066	0.098	0.022
III	0.166	0.187	0.571	0.285	0.294	1.0	1.0	0.045	0.010	0.019
IV	0.166	0.727	0.133	0.222	[a]	2.0	0.40	0.055	0.085	3.0
V	0.153	0.111	0.333	1.333	0.411	0.350	1.750	0.061	0.102	0.181
VI	0.184	0.233	0.70	0.152	0.916	0.478	5.50	0.072	0.106	0.333
VII	0.081	0.375	0.444	0.263	1.0	1.50	0.50	0.053	0.022	0.250
VIII	0.125	0.137	2.50	0.104	8.0	0.375	0.555	0.038	0.063	0.016

[a] The difference was zero.

TABLE XXII—COEFFICIENTS OF VARIABILITY OF INDIVIDUAL TESTS BY GRADES FOR VIRGINIA DELTA I

Test	Coefficients of Variability of Differences between Scores for Indicated Grades					
	3 & 4	4 & 5	5 & 6	6 & 7	7 & 8	3 & 8
I	0.111	0.266	0.129	0.111	0.190	0.033
II	0.055	0.111	0.055	0.055	[a]	0.015
III	0.333	0.166	0.20	0.285	1.0	0.181
IV	0.069	0.142	0.133	0.181	0.166	0.027
V	0.052	0.047	0.142	0.068	1.0	0.012
VI	0.048	0.105	0.062	0.073	0.214	0.018

[a] The difference was zero.

TABLE XXIII—COEFFICIENTS OF VARIABILITY OF INDIVIDUAL TESTS BY GRADES FOR WHIPPLE'S GROUP TEST

Test	Coefficients of Variability of Differences between Scores for Indicated Grades			
	4 & 5	5 & 6	6 & 7	4 & 7
I	0.116	0.113	0.285	0.055
II	0.50	0.166	0.50	0.166
III	0.50	0.227	0.206	0.114
IV-Pt. 1	0.212	0.097	4.0	0.082
Pt. 2	0.714	0.138	0.156	0.071
Pt. 3	0.454	0.267	0.30	0.125
VI [a]	1.0	0.250	0.363	0.115

[a] Test V was not used.

TABLE XXIV—COEFFICIENTS OF VARIABILITY OF INDIVIDUAL TESTS BY GRADES FOR PRIMER SCALE

Test	Coefficients of Variability of Differences between Scores for Indicated Grades		
	1 & 2	2 & 3	1 & 3
I	0.103	0.150	0.081
II	0.125	0.136	0.065
III	0.111	0.30	0.086
IV	0.081	0.115	0.047

TABLE XXV—COEFFICIENTS OF VARIABILITY FOR SERIES G AND H BY GRADES FOR VOCABULARY SCALE

Grades	Series G	Series H
3 & 4	0.072	0.089
4 & 5	0.083	0.250
5 & 6	0.096	0.111
6 & 7	0.166	0.133
7 & 8	0.136	0.122
8 & 9	0.272	0.208
9 & 10	0.235	0.161
10 & 11	0.555	0.555
11 & 12	2.000	0.333
3 & 12	0.019	0.019
3 & 6	0.033	0.039
6 & 9	0.041	0.031
9 & 12	0.083	0.075

An examination of these tables reveals the reason for the lack of discriminating power in the high school which is shown by the Otis, Classification, and Vocabulary scales. The coefficients of variability are large. Since these coefficients arise from dividing the probable errors of the differences between grade scores by these differences, large coefficients indicate small differences and large probable errors of these differences. In other words, they indicate little discrimination between the performance of the grades in question and the discrimination that exists is unreliable. When, therefore, relatively large coefficients are found in connection with the high-school grades, it means that the materials are not so graded that they will provide sufficient steps of difficulty for these grades. This suggests, then, that differentiation in the content of intelligence scales will be needed for different grade levels. We shall probably need as much flexibility in intelligence scales as has been found necessary in educational tests.

Many details are shown by the tables. For example, it will be noted that in Table XXI Test IV in the Classification Test does not appear to be properly graded for grades IX to XII. It may be added parenthetically that Dr. Theisen has been aware of this fact for some time, for the same point was revealed by his own distributions of the scores in this test. Thus, the tables furnish an opportunity for many comparisons. It is not felt that it is worth while to call attention to all of them. The chief irregularities in the tests have been pointed out in the preceding comments.

IV. *Correlations between Equivalent Tests*—Every person who has had even a limited experience with examinations and mental test work realizes that individuals do not always do their best when placed in a test situation. This variability has been noted many times in the literature of school marks and similar studies. But since formal tests have become rather widely used, many people have neglected this fact and have assumed that for all intents and purposes the scores made by individuals on a test are reliable to a high degree. This assumption is not always true.

This investigation afforded an opportunity for a consideration of this question. Several of the scales contained tests that could be considered equivalent. In some cases there was a divergence in form and structure of the test which might be responsible for variability, but the resemblance was close enough to make it seem fair to make the comparison. Accordingly, the correlations presented in Table XXVI were computed.

TABLE XXVI—CORRELATION BETWEEN EQUIVALENT TESTS

Test	Grade					
	III	IV	V	VI	VII	VIII
Arithmetic in						
Otis and Classification						0.72±0.05
Otis and Virginia				0.52±0.06	0.53±0.12	0.51±0.07
Classification and Virginia			0.58±0.05	0.35±0.07	0.43±0.07	0.70±0.05
Trabue in						
Classification and Group			0.36±0.06			
Analogies in						
Classification and Otis						0.11±0.10
Opposites in						
Otis and Classification						0.36±0.11
Otis and Virginia				0.19±0.11	0.56±0.11	0.28±0.08
Classification and Virginia			0.45±0.06	0.34±0.07	0.47±0.06	0.46±0.06
Vocabulary—Series G & H	0.41±0.05	0.20±0.04	0.38±0.04	0.40±0.04	0.58±0.04	0.55±0.04
Information in						
Virginia and Classification			0.58±0.05	0.54±0.05	0.62±0.06	0.39±0.07
Common Sense in						
Virginia and Classification			0.51±0.06	0.30±0.06	0.39±0.08	0.14±0.08

A casual inspection of this table shows that there is a great difference in the degree of reliability between comparable individual tests. Unless the repetition of a test produces approximately the same score for individuals when taken on successive occasions, its reliability may be seriously questioned. When individual tests do not show a fairly high degree of reliability as measured by the correlation between the individual scores made on the two trials, one cannot expect a scale composed of these unreliable individual tests to be very satisfactory.

It is possible that tests differing in structure and content (e.g., Arithmetic and Analogies) may measure the same mental functions. The question then arises as to whether it is better to give different tests or merely to use different forms of the same test. If the same test is repeated without rest periods, the results may not be as satisfactory as if different tests are used, because children may not concentrate as readily on the later repetitions of the test. On the other hand, later repetitions of a test might secure better results, if novelty was a distracting factor during the first attempt. For variety's sake—and variety is the spice of schoolroom testing—it may be desirable to use tests different in character. In all probability, if the scale is to be repeated, the best plan will be to administer the different forms on successive days. By this procedure the chance physiological variations which no doubt affect the child's efficiency will tend to be neutralized; for the average of several performances measured on different days usually gives more accurate data about a pupil than a single measure. This point is one that lends itself to further study and is one that ought to be made the center of careful research in order that we may have an adequate foundation for both intelligence and educational testing.

This point may be considered from another angle. Brown[6] has given us a formula by which we can determine the number of repetitions of equivalent tests needed to secure a reliability of any desired value. With this formula the number

---

[6] Brown, Wm. *Mental Measurement*, Cambridge University Press, 1911, pp. 101-102. Brown's formula for computing the reliability of the amalgamation of the scores of similar tests is

$$r_n = \frac{nr_1}{1+(n-1)r_1}; \text{ whence } n = \frac{r_n(r_1-1)}{r_1(r_n-1)}$$

of repetitions of a test needed to secure a reliability of 0.80, 0.90, and 0.95 have been computed and are presented in Table XXVII. From this table we may readily conclude that equivalent tests which do not reveal a correlation with one another of at least 0.50 are of doubtful value. Applying the values in this table to the correlation coefficients presented for equivalent tests in Table XXVI, it can be seen that a large number of the individual tests are of poor reliability if they are to be used to decide the fate of individuals. When tests are used to measure groups they are more reliable. This conclusion is supported by the only correlation coefficient which could be computed from these data—a value of 0.74±0.05, representing the correlation between the room scores in the elementary grades for Series G and Series H, Sentence Vocabulary Scale.

TABLE XXVII. NUMBER OF REPETITIONS OF A TEST NEEDED TO SECURE RELIABILITY INDICATED [a]

Correlation Coefficient	Reliability		
	0.80	0.90	0.95
0.30	10	21	45
0.35	8	17	36
0.40	6	14	29
0.45	5	11	24
0.50	4	9	19
0.55	4	8	16
0.60	3	6	13
0.65	3	5	11
0.70	2	4	9
0.75	2	3	7
0.80		3	5
0.85		2	4
0.90			3

[a] In some cases the number of repetitions given will result in a reliability slightly better than that indicated. Round numbers alone are presented because it was felt to be absurd to speak of fractional repetitions.

V. *Zero Scores in Individual Tests*—Zero scores result from tests which are not suited to the capacities of the child. Whenever such scores become appreciable they imply that the test is being administered at a point beneath its lower level of efficiency. The percents of zero scores for the individual tests in the different scales are presented in Table XXVIII. Attention is called to those which have especially high percents of such scores. They are tests IV and IX, Otis Scale; tests III and VII, Classification Test; tests IV and V, Virginia Delta I; and

TABLE XXVIII. PERCENT OF ZERO SCORES IN INDIVIDUAL TESTS FOR DIFFERENT SCALES

Scale or Test	Percent of Zero Scores in Indicated Grades							
	I	II	III	IV	V	VI	VII	VIII
OTIS								
I Following Directions	......	......	......	......	......	0	0	0
II Opposites	......	......	......	......	......	0	0	0
III Disarranged Sentences	......	......	......	......	......	2	0	0
IV Proverbs	......	......	......	......	......	8	11	11
V Arithmetic	......	......	......	......	......	0	0	0
VI Geometric Figures	......	......	......	......	......	0	0	0
VII Analogies	......	......	......	......	......	0	0	0
VIII Similarities	......	......	......	......	......	0	0	0
IX Narrative Completion	......	......	......	......	......	8	11	0
X Memory	......	......	......	......	......	0	0	0
CLASSIFICATION								
I Following Directions	......	......	......	......	0	0	0	0
II Synonym-Antonym	......	......	......	......	0	0	0	0
III Arithmetic	......	......	......	......	8	0	0	0
IV Common Sense	......	......	......	......	1	0	0	0
V Completion	......	......	......	......	4	0	0	0
VI Analogies	......	......	......	......	1	0	0	0
VII Number Completion	......	......	......	......	40	8	13	17
VIII Information	......	......	......	......	1	0	0	0
VIRGINIA DELTA I								
I True-False	......	......	1	0	0	0	0	0
II Arithmetic	......	......	6	0	0	0	0	0
III Picture Completion	......	......	0	0	0	0	0	0
IV Synonym-Antonym	......	......	25	18	6	55	2	0
V Common Sense	......	......	17	5	1	1	0	0
VI Information	......	......	8	0	0	0	0	0
WHIPPLE								
I Arithmetic	......	......	......	0	0	0	0	
II Completion	......	......	......	0	0	0	0	
III Substitution	......	......	......	4	6	0	3	
IV Reasoning Tests—								
Part I	......	......	......	0	0	0	0	
Part II	......	......	......	65	48	30	3	
Part III	......	......	......	40	30	18	9	
V Punched Hole Test	......	......	......					
VI Proverbs	......	......	......	16	23	6	0	
PRIMER								
I Dot Pattern	3	0	2					
II Classification	1	1	1					
III Form Board	1	2	0					
IV Absurdities	3	1	0					
VOCABULARY								
Series G	......	......	10	0	0			
" H	......	......	14	1.5	0			

Test IV, Parts II and III, Whipple's Group Test. There is no apparent agreement in respect to zero scores among tests of the same kind occurring in different scales. Those which do not qualify in some scales apparently do so in others. The difference in efficiency of these various tests can be attributed to their degree of difficulty and the way they are administered. The general information presented by the Vocabulary and Virginia Delta I scales seems to suggest that literary scales have a low degree of efficiency in the third grade. There are many children in this grade who have not sufficiently mastered the mechanics of literary tests to permit the measurement of their intelligence by these instruments. A zero score on a test does not imply zero mentality. It does imply an inferior mentality or lack of comprehension. In general, it may be said that zero scores when they are appreciably numerous indicate the approximate lower limits of the efficiency of a test. If a scale contains several tests which result in zero scores in the same grade, it is clear that the resulting total score cannot justly be compared with total scores wherein all of the tests function. What effect this point would have on the discriminative power of a scale is not answered by these data. It raises a question which might well be given careful study.

The question of zero scores should also be considered from the administrative point of view. Whenever a test is given which the child does not comprehend because of incomplete instruction or difficult subject matter, the problem of copying is introduced. Nearly every pupil is accustomed to take cues from others when anything is not clear. Consequently, when children in groups are asked to do something which is beyond them, they look around to see what others are doing. This behavior is not such dishonesty as would warrant a severe penalty. However, it is an argument against the administration of tests which result in many zero scores.

F—GENERAL COMPARISON OF SCALES

The foregoing data in Part II furnish many items of information concerning the six intelligence scales considered. Isolated, these facts do not tell us very much; but when summarized they prove more valuable. It will be worth our while to spend a few moments discussing the criteria by which scales may be considered.

I. *Grades for which the Scales are Suitable*—If some of the tests of a scale result in zero scores, it is not advisable to administer that scale in grades below the one in which zero scores first appear in quantity because the percent of zero scores increases very rapidly in successively lower grades and because administrative difficulties are raised when pupils are not kept busy by a test. The upper limit of a test is determined by the difficulty of the material. If total scores do not discriminate between grades, the upper limit has been reached.

II. *Ease of Administration*—Scales which have complicated directions, requiring careful study and much practice on the part of the examiner before they can be administered with any guarantee of success, are not so desirable as scales which are provided with more simple administrative procedures. Scales requiring a long time are preferable only when they are more discriminative than those requiring less time.

III. *Difficulties Involved in Scoring*—Checking the work performed by the pupils is a purely clerical task which is much the same for the different scales. Sufficient information will be furnished on this point by a mere comparison of the amounts of time required to score an individual paper.

IV. *Correlation with Scholarship*—As has already been pointed out, these correlations are rather low. Nevertheless, it is felt that the scale which correlates the highest is the best, and vice versa.

V. *Reliability*—Scales which do not reveal sufficiently large differences between the average scores of successive grades manifestly cannot discriminate between the different abilities presented by the children in those grades. When differences are accompanied by large probable errors the reliability of such differences is low—i.e. the discrimination, whatever it may be, is questionable.

VI. *Individual Tests*—Finally, the scales may be considered from the standpoint of the degree of balance shown by the different individual tests which compose them. If the individual tests vary markedly in their contributions to the total score and show inversions between averages for successive grades, the scales in which such tests occur are not as valuable as those which are well balanced and regular.

Using the criteria which we have set up in the preceding

TABLE XXIX—GENERAL COMPARISONS OF THE SIX SCALES

Scale	Range of Suitability (Grades)	Administration	Scoring	Correlation with Scholarship	Reliability of Total Scores between Successive Grades	Individual Tests
Otis	6—12	Slightly complicated and lengthy	Lengthy	0.42	poor	Erratic and poorly-balanced
Classification	5—12	Slightly complicated and of medium length	Medium length	0.50	poor	Regular but poorly-balanced
Virginia	3— 8	Simple and brief	Brief	0.59	good	Regular but poorly-balanced
Group	6— 8	Simple but lengthy	Very lengthy	[a]	[a]	Regular and well-balanced
Primer	1— 3	Simple and brief	Very brief	0.34	fair	Regular and well-balanced
Vocabulary	3—12	Very simple and brief	Brief	0.47	good	Regular and well-balanced

[a] Data not available.

paragraphs, we may compare the six scales as indicated in Table XXIX. It is not always easy to generalize, and this limitation must be kept in mind as the comparisons are examined. To some extent they are personal opinions.

G—SUMMARY AND COMMENTS

I. The present intelligence scales admit of much improvement. Some of them are erratic and poorly balanced, with comparatively poor reliability for the total scores. Others require too much time for administration or scoring. Nearly all of them can be improved by the addition of new material or by the preparation of different tests for different intelligence levels.

II. The best scales for intermediate and grammar grades seem to be the Virginia Delta I and the Vocabulary scales. A combination of these two seems to offer the best measuring instrument.

III. The Classification Test seems best for the high-school grades.

IV. The Primer Scale is well organized from the standpoint of administration, scoring, and balance, but it is of questionable diagnostic value. It should always be supplemented by other tests.

V. The value of the different intelligence scales should not be determined alone on statistical grounds. There are so many factors that influence the work of the child that it is impossible in every instance to forecast his performance in school work accurately by means of intelligence tests. There will be exceptions in most classes. These exceptions should be analyzed in the light of the information available and, in this way, many apparent deviations between scholarship and intelligence ratings may be satisfactorily explained.

VI. The present outlook for the derivation and use of group intelligence scales is good. The field is fertile and there is every indication that successful scales may be prepared.

PART III—MENTAL SURVEY OF THE CHAMPAIGN PUBLIC SCHOOLS

A—INTRODUCTION

The point of view taken in this portion of this bulletin is quite different from that assumed in Part II where our attention was directed constantly to the individual scales and tests. In Part III, on the contrary, we shall consider the child as the unit. Intelligence scales function only as they throw light on the mental ability of children.

For the purposes that we have in mind in Part III it will be most convenient to make the assumption that for practical purposes the tests which have been used are fairly satisfactory. Of course, in the light of what has been pointed out in Part II, an assumption of this sort with respect to some of the test scores would not be warranted if it were necessary to make individual recommendations on the basis of these values. This procedure, however, will give us a basis of treatment which may be copied by superintendents and teachers who wish to analyze the intelligence situations in their schools. With this point of view in mind, the statistical treatment in this section will be very simple. Everything will be determined on the basis of distributions, with the median as the measure of central tendency.

B—VARIATIONS BETWEEN SCHOOLS AND GRADES

Teachers often remark that the pupils whom they teach in one semester are not equal in mental ability to those whom they taught in some particular preceding semester. Sometimes educators are inclined to charge these opinions on the part of the teachers to personal idiosyncracies, assuming that all school grades are much alike in character. This assumption is seldom justified. Classes vary from year to year in their composition. One may have a large number of mediocre pupils in it, with comparatively few bright or dull ones. Another may have the extremes with few average pupils. In other words, classes vary considerably when the finer points are considered. On the other hand, teachers often erroneously assume that there is an intelligence difference between two classes. They are often led to make this assumption by the responses of a few superior or inferior pupils. A few exceptionally bright or dull pupils in a room will influence the tone of a group in a manner out of all proportion to their number.

I. *Variations as Shown by Median Scores in Intelligence Scales*—The variation between grades in the different schools may be shown by the differences between median scores made by the children of those grades in the different intelligence scales. These median scores are presented in Tables XXX and XXXI.

An examination of Tables XXX and XXXI shows that the same grades in the different schools are much alike in general intelligence when considered from the standpoint of the class medians. The median scores show some variation, a part of which may be due to their unreliability as measures of the group intelligence, and a part of which may be due to real differences in the grades tested. In general, however, there is a surprising uniformity and the differences which appear consistently are the ones which have been recognized by the teachers and supervisors. School No. 2 is in the best section of the city of Champaign, Illinois, and School No. 5 draws partly from the poorest section of the city. The differences between the median scores for the grades of these two schools show a decided superiority in both the Vocabulary and Virginia scales in favor of School No. 2 in every grade. If we examine School No. 9 (Table XXXI), we find that the differences between rooms are very noticeable. The pupils in the six rooms in the eighth grade had been classified into sections according to their scholarship records. At the beginning of the school year their records in the seventh grade had been used as a rough basis for classification. When the year had progressed far enough to give examinations, the pupils had been tested very carefully, and a reclassification had been made on the basis of these examination records. As a consequence, Room No. 4 contained mainly the superior pupils, while Room No. 3 contained the poorest section. The classification of the pupils on the basis of scholarship resulted in some overlapping between the different sections. The teachers recognized this fact, since they rated some of the children in Room No. 3 as of average or above average ability in scholarship. The ranks of the different sections based on scores in the Vocabulary and the Virginia scales agree closely with the ranks previously given these sections by the school officials.

TABLE XXX—SCHOOL SCORES FOR EACH SCALE BY GRADES

School and Scale	Grade						
	I	II	III	IV	V	VI	VII
School No. 1							
Vocabulary	....	....	22	29	38	42	47
Virginia	....	....	59	72	86	96	111
Primer	50	62	72				
School No. 2							
Vocabulary	....	....	20	33	31	43	48
Virginia	....	....	....	74	89	105	115
Primer	48	53	56				
Classification	....	....	....	....	59	83	97
School No. 3							
Vocabulary	....	....	....	30	36	38	47
Virginia	....	....	....	73	79	100	118
Primer	40	54	65				
Otis	....	....	....	....	....	94	117
School No. 4							
Vocabulary	....	....	14	32	34	42	
Virginia	....	....	44	63	81	92	
Primer	43	53	60				
School No. 5							
Vocabulary	....	....	....	28	30	38	44
Virginia	....	....	40	72	74	88	105
Primer	41	54	62				
Classification	....	....	....	....	58	81	100
School No. 6							
Vocabulary	....	....	8	30	31	30	45
Virginia	....	....	....	65	79	89	107
Primer	46	54	66				
School No. 7							
Vocabulary	....	....	17	28			
Virginia	....	....	48	58			
Primer	37	55	53				
School No. 8							
Vocabulary	....	....	17	24			
Virginia	....	....	43	66			
Primer	41	62	70				

II. *Quartile Variation of Children between Schools by Grades*—The preceding section, which considered merely the median scores for the different rooms, gave a general picture of the situation in the city. A general picture, however, is not entirely adequate. Two fourth-grade classes may have identical median scores but, when the individual children are considered, there may be important differences. One may have children all of whom are of approximately the same mental ability, the

TABLE XXXI—SCORES IN SCHOOL NO. 9 FOR EACH SCALE BY ROOMS

Grade and Scale	Room					
	1	2	3	4	5	6
Seventh Grade						
Vocabulary	45	43				
Virginia	111	106				
Eighth Grade						
Vocabulary	56	55	43	63	51	46
Virginia	123	120	104	136	113	119
Classification	....	....	....	116	103	
Otis	....	....	....	143		121

other may have a number of dull children balanced by a corresponding number of bright children. Manifestly, it will be much easier to teach the first class than the second. No statistical presentation can take the place of the detail furnished by the distribution table. It can be used, however, to best advantage only when a few groups of children are under observation. Under the present circumstances some general expression of variation is needed—something that will give an average "scatter" just as the median gives an average score. We shall use the quartile deviation for this purpose.

First we may regard the pupils of each grade as constituting a single group. Table XXXII shows the median and quartile deviation (half the scale distance between the 25- and 75-percentile) for each test and for each grade.

Second, we may consider the deviation of each pupil from the median of his grade. Clearly if we wish to handle these deviations together we must express them in a common or at least a comparable unit. We propose to use the quartile deviations given in Table XXXII as such units. If, for example, a fourth-grade pupil scored 43 in the Vocabulary Scale, his deviation from the fourth-grade median for that test (31) would be +12. Since the corresponding quartile deviation is 6, his deviation in terms of the quartile deviation would be +2. Similarly if a sixth-grade pupil scored 76 in the Virginia Scale, his deviation in terms of the units of the scale would be −22. One might think that this deviation is greater than that of the fourth-grade pupil just mentioned. It is true that numerically and in terms of scale units it is nearly twice as great. But the variability of sixth-grade scores on the Virginia Scale is in general greater, amounting, according to Table XXXII, to 11. In other words, the expectancy of deviation from the median is greater. This makes the larger numerical deviation of 22 of less significance. As before we may express the relation between the individual deviation and the general measure of dispersion for the group by dividing the former by the latter. When this is done we find that, in terms of the quartile deviation for the grade and test, the individual deviation for the sixth-grade pupil whom we are considering would be −2.

TABLE XXXII. MEDIANS AND QUARTILE DEVIATIONS

Scale	Grade							
	I	II	III	IV	V	VI	VII	VIII
Vocabulary								
Median	....	....	21	31	33	40	47	53
Quartile Deviation	....	....	7	6	6	6	7	7
Virginia								
Median	....	....	48	68	80	98	110	117
Quartile Deviation	....	....	10	10	11	11	12	11
Classification								
Median	....	....	....	....	59	83	99	110
Quartile Deviation	....	....	....	....	12	10	12	11
Otis								
Median	....	....	....	....	....	94	117	128
Quartile Deviation	....	....	....	....	....	10	16	13
Picture Completion								
Median	19	24	26	28				
Quartile Deviation	4	4	3	2				
Primer								
Median	44	57	64					
Quartile Deviation	9	7	7					

Table XXXIII gives the distribution of pupils of each grade according to the deviations of their scores from the median, deviations being expressed in terms of the quartile deviation for the grade to which the pupil belonged and for the test which he took. If the same pupil took more than one test the average of his deviations is given. Thus each child is entered but once. The medians shown in Table XXXIII are the medians of deviations.

These medians show in a general way the differences between the performances of pupils in the several grades. The differences in the composition of each grade of the various

TABLE XXXIII—DISTRIBUTION OF PUPILS IN EACH GRADE AND SCHOOL ACCORDING TO THEIR DISPERSION FROM THE MEDIAN. THE UNIT IS THE QUARTILE DEVIATION

Grade	Deviation $\left(\frac{x}{Q}\right)$	School No. 1	School No. 2	School No. 3	School No. 4	School No. 5	School No. 6	School No. 7	School No. 8
I	+3.0 to +3.9	1							
	+2.0 to +2.9	3			3				
	+1.0 to +1.9	3	2	3	7	2	2	1	2
	0 to +0.9	18	10	5	15	5	9	4	12
	−1.0 to −0.1	6	2	16	9	13	5	1	10
	−2.0 to −1.1	3	1	7	8	5	1	1	5
	−3.0 to −2.1		4	2	1	4		2	1
	−4.0 to −3.1					1		1	
	Total	34	19	33	43	30	17	10	30
	Median	+0.44	+0.25	−0.53	+0.25	−0.62	+0.28	0	−0.10
II	+3.0 to +3.9								
	+2.0 to +2.9		1						
	+1.0 to +1.9	11	4	2	1	3	2	4	4
	0 to +0.9	16	15	13	13	7	9	6	6
	−1.0 to −0.1	7	13	15	15	10	0	2	9
	−2.0 to −1.1	1	2	1	1	2	3	2	7
	−3.0 to −2.1	2	1	2		1	3	1	2
	−4.0 to −3.1			2		2	1	4	
	−5.0 to −4.1							1	
	Total	37	36	35	30	25	18	20	28
	Median	+0.53	+0.13	−0.17	−0.07	−0.25	+0.22	0	−0.44
III	+3.0 to +3.9	1							
	+2.0 to +2.9	3	1	1					
	+1.0 to +1.9	7	3	5	1	2	1	1	2
	0.0 to +0.9	14	11	19	8	12	6	7	12
	−1.0 to −0.1	9	10		14	7	3	6	11

TABLE XXXIII—(*Continued*)

Grade	Deviation $\left(\frac{x}{Q}\right)$	School No. 1	School No. 2	School No. 3	School No. 4	School No. 5	School No. 6	School No. 7	School No. 8
	−2.0 to −1.1	2	5	10	6	6	2	1	4
	−3.0 to −2.1		2	1	4	1	3	3	
	−4.0 to −3.1		2		1	2		1	
	−5.0 to −4.1							1	
	Total	36	34	36	34	30	15	20	29
	Median	+0.50	−0.20	+0.37	−0.57	−0.17	−0.14	−0.33	−0.05
IV	+4.0 to +4.9	1	3						
	+3.0 to +3.9	0	1						
	+2.0 to +2.9	2	3		1				1
	+1.0 to +1.9	3	5	5	3	1	2		0
	0 to +0.9	12	9	12	16	14	2		2
	−1.0 to −0.1	16	8	6	7	8	6		9
	−2.0 to −1.1	0	4	6	7	1	2		2
	−3.0 to −2.1	2	4	7		1	1		2
	−4.0 to −3.1					1			1
	Total	36	37	36	34	26	13		17
	Median	0	+0.28	−0.17	+0.16	+0.14	−0.42		−0.61
V	+3.0 to +3.9	1	1				1		
	+2.0 to +2.9	3	1	3	2	2	1		
	+1.0 to +1.9	7	10	2	4	5	1		
	0 to +0.9	7	9	13	23	12	9		
	−1.0 to −0.1	12	8	9	17	24	6		
	−2.0 to −1.1	1	3	3	9	8	3		
	−3.0 to −2.1		3			4	3		
	−4.0 to −3.1		1				1		
	Total	31	36	30	55	55	25		
	Median	+0.36	+0.33	+0.23	+0.07	−0.35	−0.08		

TABLE XXXIII—(*Continued*)

Grade	Deviation $\left(\frac{x}{Q}\right)$	School No. 1	School No. 2	School No. 3	School No. 4	School No. 5	School No. 6	School No. 7	School No. 8
VI	+3.0 to +3.9		1	2					
	+2.0 to +2.9	3	6	4		1			
	+1.0 to +1.9	3	4	4	5	4	1		
	0 to +0.9	13	14	16	20	7	2		
	−1.0 to −0.1	5	8	16	19	13	6		
	−2.0 to −1.1	5	4	6	12	8	6		
	−3.0 to −2.1	2	1	6	1	0	0		
	−4.0 to −3.1	1				1	1		
	Total	32	38	54	57	34	16		
	Median	+0.23	+0.43	−0.06	−0.18	−0.38	−0.83		
VII	+2.0 to +2.9	4	4	2					
	+1.0 to +1.9	2	5	2		3	3		
	0 to +0.9	8	13	6		10	4		
	−1.0 to −0.1	12	11	6		13	3		
	−2.0 to −1.1	4	4	3		4	4		
	−3.0 to −2.1	1	1	1		3	2		
	−4.0 to −3.1	1							
	Total	32	38	20		33	16		
	Median	−0.17	+0.23	0		−0.27	−0.33		

schools is made evident by the distributions. Some schools show a large number of below-average children, while others show the opposite condition. School No. 9, where the eighth grade was divided into sections based on scholarship achievement, shows (Table XXXIV) the decided superiority of Room No. 4 and the marked inferiority of Room No. 3. The overlapping between these rooms indicates the insufficiency of the scholarship basis when used in the classification of children into sections. The quartile deviation provides a convenient device for showing the intelligence composition of rooms and grades.

TABLE XXXIV. DISTRIBUTION OF PUPILS IN SCHOOL NO. 9 BY ROOMS AND ACCORDING TO THEIR DISPERSION FROM THE MEDIAN. THE UNIT IS THE QUARTILE DEVIATION

Grade	Deviation ($\frac{x}{Q}$)	Room 1	Room 2	Room 3	Room 4	Room 5	Room 6
VII	+1.0 to +1.9	4	4				
	0 to +0.9	8	7				
	−1.0 to −0.1	9	9				
	−2.0 to −1.1	5	7				
	−3.0 to −2.1	2	1				
	−4.0 to −3.1	1	1				
	Total	29	29				
	Median	−0.28	−0.39				
VIII	+2.0 to +2.9	2	3		4	1	
	+1.0 to +1.9	7	10	3	12	3	3
	0 to +0.9	14	7	1	10	7	9
	−1.0 to −0.1	6	11	5	2	9	8
	−2.0 to −1.1	4	2	16	3	8	11
	−3.0 to −2.1			3		2	2
	Total	33	33	28	31	30	33
	Median	+0.46	+0.50	−1.31	+1.04	−0.44	−0.56

III. *Classification of Children on the Basis of the Intelligence Quotient.*

a. The quartile deviation of children takes into account only their present intelligence status—It is almost a common opinion that two people may be rated equal in intelligence, although they may be widely different in their ability to profit by their future experience. We may make this clear by a concrete illustration. Clarence S. and Frank V. are in the fourth grade and both secure the same scores when given intelligence tests.

Clarence is 16 years old chronologically, while Frank is only 9. Manifestly, these two pupils have vastly different educational prospects. The older boy probably has reached his intellectual maturity and will soon drop out of school. He is a retarded pupil with an intelligence quotient of approximately 0.75. On the other hand, the younger child is of superior ability with an intelligence quotient of 1.10. His future educational prospects are bright. He will complete the elementary school; and if he enters high school and college, he will probably succeed. Clarence is the freight train that has started years before, while Frank is the express that has overtaken him in his educational journey.

Before it is possible to convert the scores in the intelligence tests into mental age values, it becomes necessary to establish standards for the different scales. In doing this it is necessary to make the commonly used assumption that the average individual reaches his intellectual maturity at the age of sixteen. Further, it is assumed that the highest intellectual development is represented by nineteen years' mentality. With these hypothetical bases, it is a comparatively easy matter to set up age standards for the different scales. Although the data available were incomplete at the upper and lower ranges, it is felt that an adequate allowance was made for the selected nature of the groups from which the standards were obtained. The standards used in determining the intelligence quotients of the children are presented in Table XXXV for the different scales. Below the score for each age group are two figures in italics. With these figures it is possible to interpolate ages in years and months. These standards may be used as follows: If a child makes a score of 42 in the Vocabulary Scale, he has an approximate mental age of 13.3 years. If his score is 100 in the Virginia Scale, his mental age is 13.4 years.

b. Distributions of children by intelligence quotients—The standards presented in Table XXXV made it possible to determine the mental age of each child as revealed by the scores that he made in the tests. Here again it was felt that the average of several figures is a better index than any one of them; hence the mental ages given by the various test scores were averaged for each of the children. This average mental age was divided by the chronological age, giving the intelligence quotient. In the high school where the children were older than 16 years chronologically, it was necessary to use 16 years as the divisor, since

TABLE XXXV. STANDARD SCORES FOR INDICATED AGES

Scale	4 yrs. to 4.9 yrs.	5.0–	6.0–	7.0–	8.0–	9.0–	10.0–	11.0–	12.0–	13.0–	14.0–	15.0–	16.0–	17.0–	18.0–	19.0–
Vocabulary					5 *0-10*	15 *11-18*	22 *19-25*	29 *26-32*	36 *33-39*	43 *40-46*	50 *47-53*	57 *54-60*	65 *61-69*	74 *70-78*	82 *79-86*	90 *87-93*
Virginia				15 *0-25*	36 *26-44*	53 *45-59*	65 *60-70*	76 *71-81*	87 *82-93*	99 *94-107*	115 *108-123*	129 *124-137*	145 *138-155*			
Classification						15 *0-25*	35 *26-42*	50 *43-57*	65 *58-71*	78 *72-85*	93 *86-99*	106 *100-114*	122 *115-132*	140 *133-152*	163 *153-174*	186 *175-197*
Otis							40 *20-52*	65 *53-74*	85 *75-93*	102 *94-108*	118 *109-126*	135 *127-141*	148 *142-155*	162 *156-169*	176 *170-183*	190 *184-195*
Primer	6 *0-12*	20 *13-26*	32 *27-37*	43 *38-48*	54 *49-60*	66 *61-72*	78 *73-83*	86 *84-89*								

TABLE XXXVI. DISTRIBUTION OF INTELLIGENCE QUOTIENTS BY SCHOOLS AND GRADES

Grade	I. Q.	School No. 1	School No. 2	School No. 3	School No. 4	School No. 5	School No. 6	School No. 7	School No. 8	Total
I	0.50–0.59							3		3
	0.60–0.69					2		0	1	3
	0.70–0.79			1		2	1	2	0	6
	0.80–0.89		3	2		4	0	0	0	9
	0.90–0.99		2	5		2	1	0	3	13
	1.00–1.09		4	7		5	6	1	4	27
	1.10–1.19		8	3		6	2	3	2	24
	1.20–1.29		5	1		3	5	1	1	16
	1.30–1.39		2	3			1	1	1	8
	1.40–1.49		1							1
	Total		25	22		24	16	11	12	110
	Median		1.14	1.04		1.04	1.10	1.05	1.05	1.08
II	0.50–0.59					2				2
	0.60–0.69			1		1		1		3
	0.70–0.79		1	0		1	1	1		4
	0.80–0.89		0	2	3	1	3	2		11
	0.90–0.99		3	9	4	2	1	1		20
	1.00–1.09		5	10	10	4	2	2		33
	1.10–1.19		12	6	3	3	5	3		32
	1.20–1.29		5	8	2		3	1		19
	1.30–1.39		1	1				1		3
	1.40–1.49							1		1
	Total		27	37	22	14	15	13		128
	Median		1.14	1.07	1.04	1.00	1.11	1.08		1.07
III	0.60–0.69			1	1	1		1		4
	0.70–0.79	1	1	1	1	1	1	4	1	11
	0.80–0.89	3	1	1	6	2	1	2	5	21
	0.90–0.99	5	6	6	9	6	2	4	7	45
	1.00–1.09	6	8	8	12	9	2	3	7	55
	1.10–1.19	6	10	7	3	5	2	3	7	43
	1.20–1.29	6	4	3	1	2	1	2	1	20
	1.30–1.39	3	2		1	1	1	1		9
	1.40–1.49		2							2
	Total	30	34	27	34	27	10	20	28	210
	Median	1.10	1.11	1.06	1.00	1.04	1.05	0.98	1.01	1.04
IV	0.60–0.69					1		1	1	3
	0.70–0.79	1		2	2	2		2	1	10
	0.80–0.89	2	1	5	5	6		4	3	26
	0.90–0.99	5	4	5	7	6	1	4	7	39
	1.00–1.09	6	8	7	8	5	3	2	5	44
	1.10–1.19	7	8	8	6	5	1	2	2	39
	1.20–1.29	5	8	5	4	1	2	1	1	27
	1.30–1.39	2	2	4	2			1	2	13
	1.40–1.49	2	4	1						7
	1.50–1.59	1	2							3
	Total	31	37	37	34	26	7	17	22	211
	Median	1.12	1.17	1.09	1.04	0.97	1.08	0.93	0.99	1.06

TABLE XXXVI—(*Continued*)

Grade	I. Q.	School No. 1	School No. 2	School No. 3	School No. 4	School No. 5	School No. 6	School No. 7	School No. 8	Total
V	0.50–0.59		1	1		1	1			4
	0.60–0.69	1	1	1		3	1			7
	0.70–0.79	2	0	3	3	7	2			17
	0.80–0.89	3	3	5	7	8	3			29
	0.90–0.99	4	5	9	9	7	4			38
	1.00–1.09	7	7	9	13	8	4			48
	1.10–1.19	7	9	5	7	6	5			39
	1.20–1.29	4	7	3	2	2	3			21
	1.30–1.39	2	2	3	2	2	2			13
	1.40–1.49	1	1			2				4
	Total	31	36	39	43	46	25			220
	Median	1.08	1.11	1.01	1.02	0.96	1.04			1.03
VI	0.60–0.69	2		1	1		1			5
	0.70–0.79	3	1	2	6		1			13
	0.80–0.89	4	2	9	8	4	2			29
	0.90–0.99	5	3	13	12	9	4			46
	1.00–1.09	7	8	11	17	9	5			57
	1.10–1.19	7	10	10	10	7	3			47
	1.20–1.29	4	7	5	4	4	1			25
	1.30–1.39	1	5	1	1	1				9
	1.40–1.49		2	1	0					3
	1.50–1.59				1					1
	Total	33	38	53	60	34	17			235
	Median	1.04	1.15	1.01	1.02	1.04	1.01			1.04
VII								School No. 9		
								Room 1	Room 2	
	0.50–0.59							1		1
	0.60–0.69	1					1	0		2
	0.70–0.79	1	1	2		2	2	6	2	16
	0.80–0.89	5	5	1		7	2	6	4	30
	0.90–0.99	7	6	5		9	2	4	9	42
	1.00–1.09	9	9	7		12	4	5	6	52
	1.10–1.19	4	11	2		2	4	5	3	31
	1.20–1.29	5	4	2		1	1	3	4	20
	1.30–1.39	1	2							3
	Total	33	38	19		33	16	30	28	197
	Median	1.03	1.08	1.02		0.98	1.03	0.95	0.99	1.01

it has been assumed that the average person reaches his mental maturity at this age. These intelligence quotients may be distributed to show the composition of the different grades in each school. It might be added that these values are not so reliable for children who scored low as they are for those who scored high. A child may score low through other factors than the lack of intelligence. Consequently, there are somewhat more children

TABLE XXXVII—DISTRIBUTION OF INTELLIGENCE QUOTIENTS IN THE EIGHTH GRADE OF SCHOOL NO. 9 BY ROOMS

I. Q.	Room Number						Total
	1	2	3	4	5	6	
0.70–0.79			2			1	3
0.80–0.89	1	1	7		3	4	16
0.90–0.99	4	3	10	2	7	10	36
1.00–1.09	11	9	6	6	10	10	52
1.10–1.19	11	12	2	12	8	5	50
1.20–1.29	4	5	1	9	3	2	24
1.30–1.39	1	3		2		1	7
Total	32	33	28	31	31	33	188
Median	1.10	1.13	0.95	1.16	1.08	1.01	1.07

TABLE XXXVIII—DISTRIBUTION OF INTELLIGENCE QUOTIENTS IN THE HIGH SCHOOL BY GRADES

I. Q.	Grade				Total
	IX	X	XI	XII	
0.70–0.79				1	1
0.80–0.89	4	3	3	0	10
0.90–0.99	40	41	23	16	120
1.00–1.09	44	40	48	33	165
1.10–1.19	43	37	41	38	159
1.20–1.29	18	8	15	17	58
1.30–1.39	8	4	2	1	15
1.40–1.49	2				2
Total	159	133	132	106	530
Median	1.08	1.06	1.08	1.11	1.08

showing an intelligence quotient equivalent to that of defective mentality than the facts in the case probably warrant. The excess, however, is not thought to be large; because, with the exception of the first two grades, the intelligence quotient is the combined value of the scores from several scales. As a rule, these scores varied but little with the very low-grade children, although the tests were given at different times, usually weeks apart.

The intelligence quotient is extremely significant. Although it has not been definitely established, the opinion seems to be that it remains approximately constant through life. Occasional exceptions seem to appear, but it is probable that the rule holds as steadily as do most rules regarding mental or physical characteristics. On the basis of the size of the intelligence quotient a number of classifications have been made. Terman's is commonly accepted. It is given in Table XXXIX.

TABLE XXXIX—IMPLICATIONS OF INTELLIGENCE QUOTIENTS

I. Q.	Classification
Above 140	"Near" genius or genius
120 to 140	Very superior intelligence
110 to 120	Superior Intelligence
90 to 110	Normal or average intelligence
80 to 90	Dullness, rarely classifiable as feeblemindedness.
70 to 80	Borderline deficiency, sometimes classifiable as dullness, often as feeblemindedness
Below 70	Definite feeblemindedness

The intelligence quotients presented in Table XXXVI to XXXVIII inclusive may be summarized into a single tabulation. This is presented in Table XL. The median of this table raises an important question. Its value is 1.06 which implies one of two things: either that the standards established for the different age groups have been set a little low or that Champaign represents a group slightly above the average in mentality. It is the opinion of the writer that the latter conclusion should be drawn.

An examination of Table XL gives considerable material for speculation. In the light of the percents shown in this distribution, it might be said that 1.9 percent of the school population in Champaign was definitely feeble-minded. Of course, a few pupils may have fallen into this group because they failed in the test through physical causes and not through lack of mental ability. But on the other hand, it might be added that there were several children in the schools who were too low in mentality to take the tests. Due to the fact that the teachers did not always give the data for these children, several of them were not included in the study. Some of them are hopeless imbeciles and probably will never learn to read or write.

If we adopt the standards of normality which are usually used, we shall find that only 46 percent of the school population is in the normal group—i.e., in the group which ranges from 0.90 to 0.110. The next higher group, 1.10 to 1.19 contains 23 percent. These are the two groups in which the greatest number of cases occur. Thus about seven out of every ten children in the Champaign schools appear to be either "normal" or "superior," using Terman's classification. Below this large central group we have approximately 15 percent of the school population; above it, almost 17 percent.

TABLE XL. INTELLIGENCE QUOTIENT DISTRIBUTIONS FOR THE ENTIRE SCHOOL POPULATION

I. Q.	No. of Pupils	Percent
0.50–0.59	10	0.5
0.60–0.69	27	1.4
0.70–0.79	81	4.0
0.80–0.89	181	8.9
0.90–0.99	399	19.7
1.00–1.09	533	26.3
1.10–1.19	464	22.9
1.20–1.29	230	11.4
1.30–1.39	81	4.0
1.40–1.49	20	1.0
1.50–1.59	4	0.2
Total	2,030	
Median	1.06	

In its school administration Champaign should make an attempt to provide for these two extreme groups. At present they are found in the regular classes. There are but few attempts to meet the special needs of these children. The problem is not so complicated as this table might lead one to believe. The poorest 15 percent are found almost entirely in the grades below the eighth. Moreover, they are much more numerous in some sections of the city than in others. This fact will make it possible to provide special classes for them in which the curriculum can be modified to meet their particular needs. This group probably should be provided with more of the vocational and less of the academic subjects. It is probable for example, that a course in dishwashing, sweeping, cleaning, and in other simple household duties would be very beneficial to the girls who have intelligence quotients below nine-tenths. An examination of Table XXXVI shows that there are a sufficient number of these children to provide classes of economical size from the administrative point of view.

The 17 percent who are above the large central group should also be specially provided for. These are the children who could make rapid progress through the school, if given the opportunity. It is decidedly unsatisfactory to give these children the same bill of fare as is provided for the average and the mediocre. In instruction one "exposure" of most topics is sufficient for them. A single reading of their lesson suffices, where the average or mediocre child must read it several times. From every point of view it is wasteful to keep these children in the same classes with the other children. It is not altogether satisfactory to give them rapid promotion from grade to grade, and thus to allow them to skip parts of the work. Consequently, the only sensible thing to do is to provide special classes for them.

Special classes, especially those for subnormal children, will no doubt meet with some opposition on the part of the parents, if the classes are established too abruptly and without the utmost tact. The present classification of the children in the city, which permits the transfer of a child from one school to another for administrative reasons, also permits a grouping of the children which will secure these desired results without any special advertising of the fact.

c. Age-grade intelligence-quotient groups—Regarding the child three facts not thus far combined in our tables are of special administrative importance. These are his chronological age, his grade in school, and his intelligence quotient. Manifestly, if a child is chronologically older than the normal age for the grade in which he is located and has an intelligence quotient that indicates approximate normality, the sensible thing to do is to promote him to the next higher grade, giving him extra attention so that he may meet the deficiencies in his scholarship preparation which result from this unusual progress. Champaign would experience no difficulty from the double promotion of such children, because the city has provided "opportunity classes" which all children who are maladjusted may attend in order that they may receive extra attention and make up work. These "opportunity classes" could provide for children who are now found to be out of place and do it without special effort.

The three facts which we have mentioned—age, grade, and intelligence quotient—may be presented in tabular form, thus furnishing a convenient means for discovering how many children may be considered improperly placed. Table XLI gives the distributions of intelligence quotients at each age for the the sixth grade and may serve as an example of this type of representation. An examination of this table shows that comparatively few retarded children have an intelligence quotient high enough to warrant their promotion. Perhaps eleven of the 13 year-olds and one of the 14 year-olds might be advanced. But when we consider the fact that the intelligence level of the Champaign system is 1.06 it may be questioned whether it is advisable to promote any of these children except the one who is 14 years old and whose intelligence quotient is more than 1.00. In other words, there is comparatively little maladjustment from this point of view and certainly no extreme maladjustment. This table, which is typical of the other grades, emphasizes the fact that we shall always have retardation in our schools as long as children are grouped in classes where all must take the same curriculum without reference to their ability. It seems to the writer that the most sensible provision which can be made is to classify the children into at least three groups based on the intelligence quotient and then to prepare courses of study suited to these groups.

TABLE XLI. AGE-INTELLIGENCE-QUOTIENT DISTRIBUTION FOR SIXTH GRADE

I. Q.	Age in Years					
	10	11	12	13	14	15
0.50—0.59						
0.60—0.69					4	2
0.70—0.79			1	8	9	3
0.80—0.89		1	10	17	5	3
0.90—0.99	1	7	26	14	2	
1.00—1.09		20	26	8	1	
1.10—1.19	1	23	15	3		
1.20—1.29	1	11	1			
1.30—1.39	1	5				
1.40—1.49		1				
1.50—1.59	1					
Total	5	68	79	50	21	8
Median	1.25	1.13	1.01	0.90	0.77	0.77

So far as ability to do higher-grade work is concerned it is not the older but the younger children who are really retarded. The reader will observe the consistently higher ranges of intelligence quotients among the younger children. The 15 twelve-year-old children whose intelligence quotients are between 1.10 and 1.19 have an average mental age of 13.8 years.

If they had received proper instruction and had been advanced according to their ability they would be in the seventh grade—some of them in the eighth. Similarly the eleven eleven-year-olds whose intelligence quotients are between 1.20 and 1.29 have mental ages which would entitle them to belong to the seventh grade. The five children of the same age who have intelligence quotients between 1.30 and 1.39 would be in the eighth grade, if their advancement had kept pace with their mental development. Whether such children should suddenly be promoted to the grade to which by mental age they belong is a debatable question. Such a belated adjustment would be at best a makeshift compared to the gradual adjustment that would have been possible if these children could have been identified early in their school career. Meanwhile, however, it is proper to point out that when pupils have been boldly promoted in accordance with their mental ages, even after they have been "discovered" relatively late in their school course, they have usually maintained themselves with credit in their advanced grades.

C—SUMMARY AND COMMENTS

I. Schools vary appreciably in the distribution of pupils according to mentality. Some of the rooms in the same school are much superior to others. These large differences might have been anticipated from the general opinions of the teachers and supervisors.

II. Each child's departure from the median of his grade, in terms of the quartile deviation, serves as a convenient means of comparing his performance with that of another child whether the latter be in the same grade or not—or indeed whether he has taken the same test or not. This measure serves also to indicate the large irregularities in the classification of the children.

III. The intelligence quotient is a better means of measuring the individual variability of children within a room because it emphasizes the educational possibilities of each child. It thus becomes the best device, from the standpoint of the teacher, for measuring the "brightness" of the children.

IV. Grouping the children in Champaign on the basis of intelligence quotients, 69 percent of the school pupils may be considered of approximately uniform ability. Above this central homogeneous mass there are 17 percent who are above average and of high ability. Below it are 15 percent who are as inferior as the others are superior. Special provisions with modified curriculums should be made for each of these two groups.

V. An age-intelligence-quotient table for each grade furnishes the best device for analyzing a school situation. This representation reveals the situation at a glance. If there are any maladjustments among the older groups, the fact is self-evident when the data are tabulated in this form. When the pupils who are out of place have been identified, it is a comparatively simple matter to apply the necessary remedies.

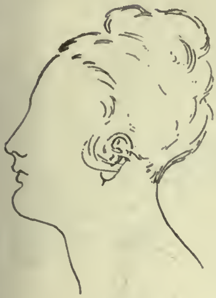
APPENDIX

I. *Incomplete Pictures as Tests*—The incomplete picture has been used consistently as a mental test since Binet introduced it in his scale. When it became necessary to devise a test for illiterates and foreigners in the army, incomplete pictures were accepted as a promising form of test. In the Beta examination devised by the army workers this form of test was used with adults of varying degrees of mental maturity, although the value of the incomplete picture as a test was comparatively unknown. Moreover, the data presented in Part II for the Virginia Delta I Scale raises the question of the suitability of the picture completion test for children beyond the primary grades because it shows almost no discriminative power for grades III to VIII. In the light of these facts it seems worth while to consider the value and possibilities of incomplete pictures as tests. Bearing upon this question are the results of a study of this form of test which has been made by the writer.

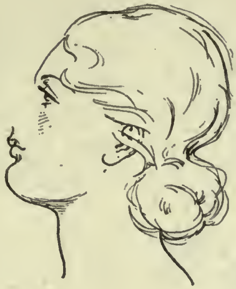
The data secured show that very few purely incomplete pictures are difficult enough to test the intelligence of normal children who are more than nine years of age. Furthermore, normal children of the primary grades are sometimes very much puzzled by incomplete pictures, although they may show a reasonable degree of keeness in other respects. In other words, the ability to recognize the omissions is a more or less specialized ability which depends to a certain extent on the type of experience the individual has had. When the problem of investigating the incomplete picture was first attacked it was fondly hoped that a series of incomplete pictures might be found which would present a range of difficulty capable of testing pupils from the primary through the grammar grades. This hope was not realized and the writer is led to the conclusion that incomplete pictures are of little value as tests of intelligence above the primary grades.

Thirty-eight incomplete pictures were submitted to approximately one thousand four hundred children. Of the total number of pictures presented twenty were found to be suitable for use in a test for the primary grades. These are presented in a "Picture Completion Test" published by the Bureau of Educational Research. Figure 1 shows this test.

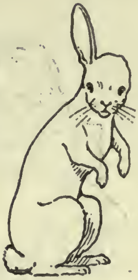
FIGURE 1. PICTURE COMPLETION TEST BY DR. CHARLES E. HOLLEY



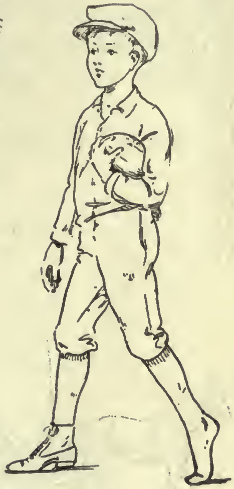
1



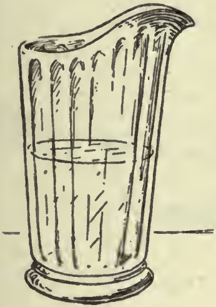
2



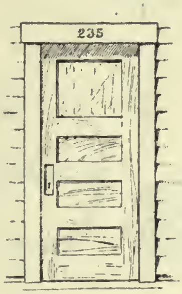
3



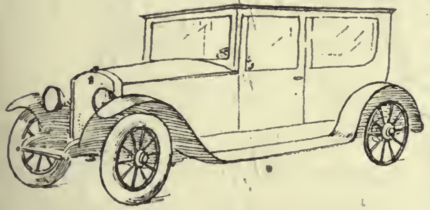
4



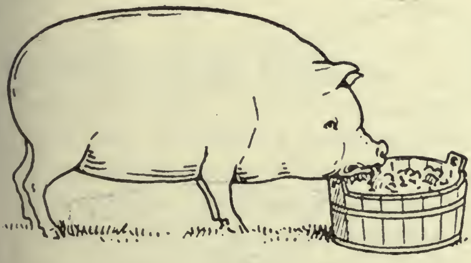
5



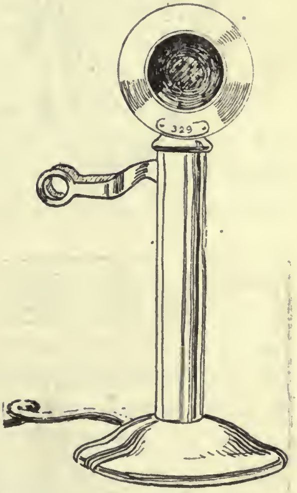
6



7

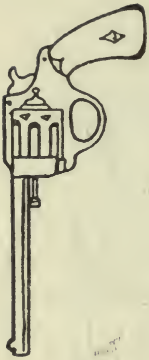


8

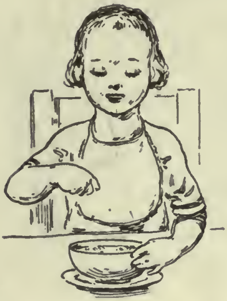


9

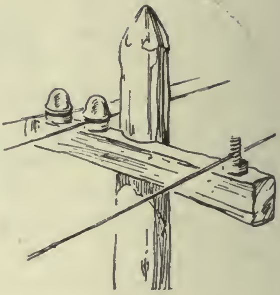
FIGURE 1—(*Continued*)



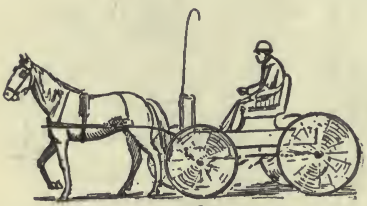
10



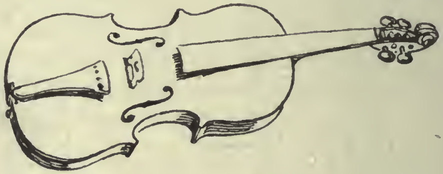
11



12



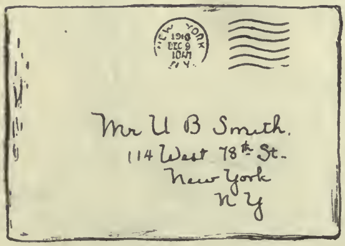
13



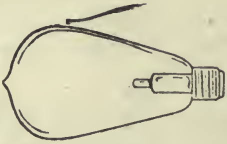
14

15

16



17



18

19

20

The difficulty in terms of the percent of correct responses was obtained for each of these twenty pictures. Table XLII shows these percents for the first three grades.

II. *Sex Differences*—The data gathered in Champaign from all the tests were examined from the standpoint of the difference between boys and girls. The results were such as to lead one to conclude that there are no real sex differences in general intelligence which may be revealed by these general tests. The medians and averages for the two sexes were approximately the same in the various tests, all differences being small enough to justify one in attributing them to chance factors.

TABLE XLII. PICTURE COMPLETION TEST. PERCENTS CORRECT FOR EACH PICTURE BY GRADES

Number	Grade		
	I	II	III
1	91	100	100
2	86	95	97
3	84	98	97
4	77	89	97
5	77	89	95
6	63	82	94
7	60	82	86
8	56	62	87
9	49	59	69
10	46	56	68
11	47	56	61
12	44	63	78
13	44	61	72
14	39	59	78
15	30	37	53
16	30	33	53
17	25	39	66
18	21	72	91
19	15	45	51
20	10	30	53

III. *Administration by Teachers*—It is highly desirable to have intelligence scales that can be administered by the individual teacher. If, however, several rooms are to be compared, it is better, in practice, for a supervisor to give the tests than for the individual teacher to do so. Although most teachers will do their best to follow instructions accurately and thus to secure uniform results, a small minority will persist in varying conditions to suit their own ends. From many points of view, it will be better if the pupils take intelligence tests under the direction of supervisors or of persons other than the room teachers. The results will be comparable from room to room and they may be made the basis of administrative measures in a way that would not be the case if their reliability were in doubt. Since nearly every test requires practice for its successful administration, the supervisor who administers a test several times becomes practiced in its details and thus secures the complete cooperation of the pupils. With the more difficult scales the supervisor can take the time to perfect his method by practicing on his friends before he administers the scale in the classroom where the results are important. Individual teachers cannot spare the time and trouble needed to perfect their technique even if they are entirely in sympathy with the work. Consequently, it is much better if all tests of this nature are administered by supervisors.

IV. *Scoring*—If mental tests are to be used in a way that will contribute most to school problems they must be scored very accurately. Where a child's future is to be influenced by the result it is vitally important that his score be as nearly correct as the test will permit.

The best results are secured if the scoring is done by a few careful workers who have been trained for their duties instead of by a larger number of people who devote only a little time individually to the work. It requires much valuable time to instruct the scorers in the methods of evaluating the various parts of the pupil's answers, and as a rule, several days of practice are needed before they can become proficient. Consequently, paid trained workers are a decided economy over volunteer workers. As far as possible nothing should be left to independent judgment. Instructions should be prepared which will cover every possible case. The system of indicating rights and wrongs on the tests should be worked out so carefully that it will economize time and eliminate chances for error. All of these details should be covered in the training of the clerical workers.

It is possible to secure good clerical workers among the student body of the average high school and college. Scoring takes good eyesight and ability to learn quickly. People who score for a few days on a test acquire increased proficiency from day to day and reach their maximum in about a week. This fact suggests that it is uneconomical to employ a large number of workers for a small project. It is better to secure a few capable workers and allow them to perform the same type of work for a longer time than would be needed for the larger group. They will become very skilled and will do the work at a much smaller unit cost.

It is best to rescore or check nearly every operation that is involved in rating a set of papers. This checking should be done by persons other than those who scored the papers the first time. A second scoring will catch most of the errors.

Stencils facilitate the work of scoring by economizing eye movements and pencil marks. Some types of material lend themselves to the use of transparent stencils. These stencils may be made from the celluloid used as window material for automobile curtains. Transparent paper may also be used—especially when durability is not necessary. Ink dots or lines may be so placed on this material that they will coincide with the marks that the pupils must make in indicating the right answers to the tests. In other types of tests cardboard stencils may be made which will enable the scorers to check the answers quickly. Every device should be employed which will economize time and insure a high degree of accuracy. Any device which leads to simplicity removes sources of error.

16

BUREAU OF EDUCATIONAL RESEARCH—BULLETIN No. 5

Report of Division of Educational Tests for '19-20

BY

WALTER S. MONROE

Assistant Director

Bureau of Educational Research
University of Illinois

PRICE, 25 CENTS

PUBLISHED BY THE UNIVERSITY OF ILLINOIS
URBANA, ILLINOIS

Bulletins of the Bureau of Educational Research

B. R. BUCKINGHAM, *Editor*

EDITORIAL INTRODUCTION

We should like to have the reader consider this monograph as, in a certain sense, "chips from the work-shop." We hold that no organization, such as that from which this bulletin emanates, should collect from users of test materials the results which they have attained in their localities and hoard them in miserly fashion for its own purposes. Moreover, it is not so nominated in the bond. It is understood that when copies of score sheets are mailed to us we are to combine them into master score sheets and to issue tabulations which will indicate over a wider field than any school system affords the conditions disclosed by the tests in question.

Although the Bureau of Educational Research of the University of Illinois went out of the business of distributing tests last November, there had been collected up to that time a valuable body of data which was augmented during the succeeding months until it now appears to justify publication.

Aside, therefore, from chapters three and four, which deal with Monroe's Standardized Reasoning Test in Arithmetic and his Timed Sentence Spelling Test, the bulletin is devoted to presenting material which will make it possible to use a number of tests more intelligently. We do not, as Dr. Monroe says, attempt in this bulletin to give directions for administering tests or interpreting results. We are mainly concerned with what the results are. We are continually receiving questions from practical workers in the field. These questions have led us to believe that they are much interested in and perplexed by the question of standards.

Realizing this fact we have tried to make our presentation of results as complete and helpful as possible. They are presented substantially in three ways. First, we show for each test the median scores by grade. In tables devoted to this sort of data we also include the 25- and 75- percentiles. To those who understand the meaning of these latter figures the nature of the distribution out of which the medians arise will be made evident. If the 75- and 25- percentiles are far apart it means that the data are scattering. In other words that the distribution of scores spreads over a wide range.

In order that there might be no doubt about the nature of the distribution, we have in the second place presented for each test, the number of pupils attaining the indicated scores in each grade to which the test was applied. The value of this sort of a showing is greater than the practical teacher is likely to realize upon the first inspection. Such a table may be converted into a table indicating a distribution in terms of percents by dividing each of the entires by the column totals. When the table is thus converted it becomes directly comparable with a similar table which may be computed for a school or school system. Moreover, since it is customary to give the grade medians in connection with this type of table—and the custom is followed in this bulletin—the teacher may learn from these figures the number and percent of pupils in each grade exceeding or falling short of the median of other grades. A teacher may likewise discover from such tables a number of subordinate facts concerning the test and its applicability to the grades in question—such facts as the number of zero scores or scores in the neighborhood of zero, the number of perfect or nearly perfect scores, and the nature of the distribution of the frequencies throughout.

But it is probable that the greatest usefulness of these distribution tables is of another sort. They are indispensible to those who wish to contribute toward the better standardization of tests. For example the 3000 pupils, more or less, in each grade, whose scores are shown in Table II for Monroe's Reasoning Test may reasonably be thought to be insufficient for a final standardization. This tabulation provides a form and makes a beginning for a more reliable treatment of the test in question. Any superintendent can place the pupils whom he has tested—be they few or many—in this scheme. Any bureau of research may gather scores from schools and school systems in this manner. After a little it may (and indeed it should) publish its findings in this manner to the end that more reliable standards may be secured. It is because tables of this sort are costly to print and of little direct school use that they are so seldom seen. They are frequently found after they have been converted into percentage distributions, because the latter are useful in making comparisons. But they are seldom found in mere frequency form. Yet the presentation of such tables is fundamental to cooperative effort. In our judgment every research organization ought to publish material in this form. Its high value for research purposes should be appreciated in contrast with its low value for immediate practical purposes.

On the other hand, the third form of tabulation is of most value for school uses. We are referring to the percentile tables presented in the appendix. We are convinced that when this type of material is better understood it will be much more widely used. By means of it a teacher may "place" a pupil's score among one hundred scores, arranged from highest to lowest, these one hundred scores being regarded as typical. Thus the percentile table will enable a fifth-grade teacher to state that a pupil is (say) twentieth among one hundred typical children of his grade in speed of reading, that he is thirty-seventh in the operations of arithmetic, fiftieth in spelling, etc. If he is fiftieth in spelling we have the special case of the median, which we ordinarily arrive at from another point of view.

In using percentile tables such as those given in the appendix of this bulletin, regard must be had for the source of the tables. In its ideal form a percentile table is supposed to have been derived from a sufficient random sampling of a total "population"—e. g., from the entire fifth-grade in American schools, or from the entire number of ten-year-olds in rural schools, or from the entire number of graduates of the Chicago high schools. In ranking a child's performance one must be sure either that he belongs to the population to which the table refers or that the population of both the child and the table are indicated. Thus if a fifth-grade pupil obtains a score in composition equal to the 80-percentile for his grade, we thereby define his rank as twentieth from the top (or eightieth from the bottom) among one hundred typical fifth grade children. A pupil thus ranked has evidently done rather well compared with pupils of his own grade. Very appropriately therefore, we may wish to rank him with reference to the sixth grade. His score may perhaps equal the 60-percentile of the sixth grade. Accordingly, he would be ranked, on his performance, as fortieth from the top among one hundred typical sixth-grade children. Similarly he may rank as fiftieth (median) for the seventh grade.

We submit these percentile tables for their practical utility. They are, however, based upon a limited number of cases; and they will be somewhat modified when more scores have been made available.

From the above statements it will be clear that the chief purpose of this bulletin is to furnish an accounting of the test results which we have received. Nevertheless, we have included two chapters (III and IV) on the derivation of Monroe's reasoning tests and timed sentence tests. These accounts have been held up a long time. When, therefore, they were released, we took account of the demand that has been made for then—especially the one relative to the reasoning test—and incorporated them into this report of the Division of Educational Tests.

TABLE OF CONTENTS

Chapter Page

I. INTRODUCTION.................... 9

II. TENTATIVE GRADE NORMS 12

  A. Monroe's Standardized Reasoning Tests in Arithmetic.... 12

  B. Buckingham's Scale for Problems in Arithmetic.................... 14

  C. Monroe's Diagnostic Tests in Arithmetic.................... 16

  D. Monroe's Standardized Silent Reading Tests.................... 19

  E. Charters' Diagnostic Language, and Language and Grammar Tests 25

  F. Willing's Scale for Measuring Written Composition.................... 26

  G. Harlan's Test for Information in American History.................... 28

  H. Sackett's Scale in United States History 29

  I. Hotz's First Year Algebra Scale 30

  J. Minnick's Geometry Tests 31

  K. Holley's Sentence Vocabulary Scale.................... 32

  L. Holley's Picture Completion Test for Primary Grades.................... 34

III. THE DERIVATION OF MONROE'S STANDARDIZED REASONING TESTS IN ARITHMETIC 36

IV. MONROE'S TIMED SENTENCE SPELLING TESTS AND PUPILS' ERRORS. 48

APPENDIX: PERCENTILE SCORES.................... 57

1. Monroe's Standardized Reasoning Tests in Arithmetic.
2. Monroe's Standardized Silent Reading Tests.
3. Charters' Diagnostic Language Tests for Grades III to VIII.
    A. Pronouns
    B. Verbs
    C. Miscellaneous A
    D. Miscellaneous B
4. Charters' Diagnostic Language and Grammar Tests for Grades VII to VIII
    A. Pronouns
    B. Verbs
    C. Miscellaneous
5. Willing's Scale for Measuring Written Composition
6. Harlan's Test for Information in American History.
7. Hotz's First Year Algebra Scale
8. Holley's Sentence Vocabulary Scale.

CHAPTER I

INTRODUCTION

Source of data. In distributing educational tests the Bureau of Educational Research has always supplied a duplicate class record sheet on which was printed a request that the duplicate be returned after the scores had been entered upon it. There has been no effort to follow up the purchasers of the tests in order to secure complete returns of the scores. Consequently, this report is based upon the scores voluntarily contributed. The bulk of the scores are from medium sized cities. Reports have been received from a few large cities (population of 100,000 or more) for Monroe's Standardized Silent Reading Tests but in no other case has more than one such city reported. Practically no scores were reported from rural schools except for Monroe's Standardized Silent Reading Tests. Several tests distributed by the Bureau of Educational Research are not included in this report for the reason that the number of scores reported seemed to be too small to justify any announcement of median scores which would be useful as tentative standards.

Form of the report. The distributions of scores entered upon the class record sheets were combined to form a total distribution of scores for each yearly grade. No attempt was made to keep separate the scores of the A and B sections of the yearly grades. In addition to the median scores for each grade the 25– and 75– percentile scores are also given for several of the tests. In a few cases the total distributions are given because they have a special significance. In the Appendix of this bulletin the 5-, 10-, 20-, 30-, 40-, 50-, 60-, 70-, 80-, 90-, and 95- percentile scores are given for a number of the tests. The publication of the percentile scores is prompted by the desire to make possible a more accurate interpretation of a pupil's score than merely that it is above or below the median score. Pupils belonging to any grade exhibit large individual differences. For this reason it is frequently desirable to know where the score of a pupil places him in the total distribution from which the median for his grade was calculated. The percentile scores make it possible to ascertain for any pupil his approximate position in this total distribution. For example, if a pupil's score is equal or superior to the 80- percentile score for his grade, he ranks in the upper 20 percent of all pupils to whom the test was given in that grade.

Time of testing. No attempt was made to organize the giving of the tests. Consequently, the scores on which the median and percentile scores are based represent testings all the way from September to June. This condition makes the derived scores somewhat less useful as tentative grade standards than they would be if they were based upon measures obtained at some one fixed time during the school year. The situation is further complicated by reason of the fact that some of the schools reporting have semi-annual promotion while others have annual promotion. Some of those which have semi-annual promotion combined the A and B divisions of each grade in making their reports. In order that the median and percentile scores shall have as definite a meaning as possible we have estimated for each test the month of the school year which the median scores appear to represent. This estimate, however, must be considered only approximate.

An organized, cooperative plan which would have resulted in the tests being given on one or more fixed dates during the school year, was not attempted for two reasons. In the first place, the complete realization of such a plan was impossible because in a large number of school systems the tests were given as a part of a local plan. The Bureau of Educational Research believes it is wise to encourage this use of educational tests. When the tests are given merely as a part of a project originated by a central bureau of research, little use is likely to be made of the results by the schools giving them. Their motive is to cooperate with the central bureau, and when this has been completed there is a tendency for them to feel that all has been accomplished which may be accomplished. This results in a great waste. Information which might be of much value to the local school systems is not used. Furthermore, this practise tends to engender an attitude toward educational tests that they are merely tools of research to be used by central research bureaus and not tools which may be used by a school system, or even a teacher, in improving instruction.

A second reason for not attempting to organize the giving of the tests was that tests were distributed in all sections of the United States and in a few foreign countries. It would have been impossible to solicit the cooperation of all who gave the tests in a plan of organized testing.

The interpretation of scores by comparison with grade norms. It is not the purpose of this bulletin to give detailed suggestions concerning the use of the grade medians and percentile scores in the interpretation of the scores obtained in any school system by giving the tests. In another place[1] the writer has indicated in some detail the general procedure to be followed in interpreting scores for the purpose of improving instruction. Grade norms are also useful in interpreting the scores of pupils for the purpose of classification.[2]

In any interpretation of scores, either individual or group, it is necessary to bear in mind certain limitations. In the first place none of our educational tests yield scores which are absolutely accurate. The errors of measurement are large in comparison with the errors made in the measurement of physical objects. Errors larger than the difference between the median scores for successive school grades frequently occur although in the case of our better tests the majority of the errors are less. These errors of measurement are chance errors and for that reason tend to neutralize each other in the median and average scores of groups. Therefore, the group scores are more accurate than individual scores. However, in interpreting either type of scores one should bear in mind the possible errors of measurement which they may include.

---

[1] Monroe, Walter S. "Improvement of Instruction Through the use of Educational Tests," *Journal of Educational Research*, I (February, 1920), 96-102.

[2] Buckingham, B.R. "Suggestions for procedure following a testing program—I, Reclassification," *Journal of Educational Research*, II (December, 1920), 787-801.

In the second place the score which a pupil makes on any subject-matter test, such as reading, arithmetic, history, or language, depends in part upon his general intelligence. Pupils belonging to any school grade differ widely with respect to their general intelligence and consequently may be expected to differ in their achievements. For this reason some pupils belonging to a given grade should have scores above the median, while others may be expected to have scores below the median because of their differences in capacity to learn. There are also differences in the average general intelligence of pupils belonging to the same school grade. For example, the average general intelligence of the fifth-grade pupils in one school may be a year or more above that of the fifth-grade pupils in another school. It is unfair to both pupils and teachers to interpret achievement scores without recognizing the differences which may exist in the general intelligence of the pupils. To do so will frequently result in arriving at erroneous conclusions. Hence, grade standards such as are given in this report must be used with due caution.

Chapters III and IV contain reports of studies which were made by the writer during the time he was Director of the Bureau of Educational Measurements and Standards of the Kansas State Normal School, Emporia, Kansas. These reports were originally prepared for publication by that institution. Permission has been obtained to incorporate them in this bulletin. In doing this the manuscript has been only slightly revised. Chapter III gives an account of the derivation of the Monroe Standardized Reasoning Tests in Arithmetic. Chapter IV contains a description of the derivation of Monroe's Timed Sentence Spelling Tests and a report of a study of pupils' errors in spelling based upon them.

CHAPTER II

TENTATIVE GRADE NORMS

The percentile scores as well as the median scores which are given in this chapter should be used only as tentative grade standards. For several of the tests the number of scores on which these are based is so small that the standards can not be thought of as final. When other scores are added to the distributions, it is likely that different medians will be obtained. Furthermore, such standards should always be thought of as representing the average of present conditions and not as being ideal standards or what ought to be.

A. Monroe's Standardized Reasoning Tests in Arithmetic

The derivation of these tests, which consist of a series of one- and two-step problems, is described in Chapter III. For each problem two values were calculated, "correct principle value," or P, and "correct answer value," or C. These values represent the credit which is to be given for solving the problem correctly in principle and for obtaining the correct answer. Each problem is marked for correct principle. If a problem is solved correctly in principle it is further marked with reference to correct answer. A pupil does not receive credit for a correct answer if the problem was solved by the wrong principle. The directions for administering the tests provide for having the pupils mark the problem on which they are working at the end of ten minutes. In this way

TABLE I. MONROE'S STANDARDIZED REASONING TESTS IN ARITHMETIC. FORM I. GRADE NORMS FOR APRIL TESTING

	Grade				
	IV	V	VI	VII	VIII
Correct Principle					
Number of pupils.	2932	3027	3498	2796	2472
25-percentile	6.2	12.1	10.0	13.8	11.5
Median	11.3	19.2	14.2	19.7	17.2
75-percentile	16.8	25.9	19.4	24.7	22.8
*Rate					
Number of pupils	1412	1705	1699	1717	1642
25-percentile	5.2	8.0	6.4	8.0	5.3
Median	7.8	11.2	8.7	11.2	7.5
75-percentile	8.1	15.1	12.1	14.5	10.9
Correct Answers					
Number of pupils	2968	2996	3518	2803	2515
25-percentile	4.1	7.1	6.9	9.4	5.1
Median	7.0	11.3	10.4	13.4	9.0
75-percentile	10.7	15.5	14.0	17.4	13.0

*Sum of correct principle values of problems done correctly within ten minutes.

a rate score may be obtained. It is the sum of the "principle values" of the problems which are solved correctly in principle within ten minutes. However, the obtaining of the rate score is optional, and it was reported in only about half of the cases.

There are two forms of these tests. These forms were constructed so that they were expected to be equivalent. Experience in using them suggests that they are not equivalent, although data are lacking at this time on which a statement concerning their comparability may be based. No scores are reported for Form 2 because the returns received for this form included an insufficient number of cases.

Test I is given in Grades IV and V, Test II in Grades VI and VII, and Test III in Grade VIII. The tests were not constructed so that the scores yielded by the different tests are comparable. Therefore, direct comparisons can not be made between the fourth and fifth grade scores and between the seventh and eighth grade scores.

TABLE II. MONROE'S STANDARDIZED REASONING TESTS IN ARITHMETIC FORM I, CORRECT PRINCIPLE

Score*	Grade				
	IV	V	VI	VII	VIII
43	3	11			
41		5			
39	7	20			
37	1	21			
35	12	89			
33	11	56			
31	26	137			56
29	25	120	47	127	94
27	40	191	93	262	63
25	80	191	131	202	171
23	89	223	135	242	214
21	124	269	248	304	233
19	130	207	280	322	214
17	161	211	306	259	217
15	248	231	328	225	237
13	260	219	470	225	203
11	298	167	425	193	201
9	267	166	374	134	133
7	294	148	276	103	154
5	304	140	191	52	121
3	230	94	123	30	87
1	185	54	49	18	51
0	137	57	22	8	23
Total	2932	3027	3498	2706	2472
Median	11.3	19.2	14.2	19.7	17.2

* In this bulletin all intervals unless otherwise noted are expressed in terms of their lower limits.

Table I gives the grade medians, 25-percentile, and 75-percentile scores for correct principle, correct answer, and rate. The distributions of scores for correct principle are given in Table II. These indicate that Test I is too difficult for a number of pupils in the fourth and fifth grades. In the construction of the tests no effort was made to include very easy problems. In fact, as is shown in Chapter III, the difficulty of a problem was not considered as a basis for selection. In none of the other grades do the zero scores amount to as much as one per cent of the total. In the seventh grade nearly five percent of the pupils made perfect scores.

REFERENCES

Willing, M. H. "The Encouragement of Individual Instruction by Means of Standardized Tests," *Journal of Educational Research*, I (March, 1920), 193-198.

Results from the Monroe Scandardized Reasoning Tests are used to illustrate how such work as the title mentions may be carried on. Suggestions for diagnosis of faults, remedial measures, etc. are given.

B. BUCKINGHAM'S SCALE FOR PROBLEMS IN ARITHMETIC.

The problems for Buckingham's scale were selected largely on the basis of diffculty. Division One is for Grades III and IV, Division Two for Grades V and VI, and Division Three for Grades VII and VIII. The problems of Division One increase by steps of approximately 0.3 P. E. from 2.7 to 5.3. The problems of Division Two increase by similar steps of difficulty from 5.5 to 7.3, and the problems of Division Three increase from 7.5 to 9.4. In scoring the test papers attention is given only to the numerical accuracy of the answers. A pupil's score is the difficulty value of the hardest problem which he answers correctly, unless he has failed on one or more previous problems. In that case, a correction is made by subtracting from the value of the hardest correctly solved problem 0.3 for each failure in Division One, or 0.2 for each failure in Division Two or Three. Thus, if a pupil solved the first six problems in Division One, his score is 4.2; but if he fails on the 4th and 5th (otherwise succeeding through the 6th), his score is 3.6—i.e., 4.2 —2 x 0.3.

TABLE III. BUCKINGHAM'S SCALE FOR PROBLEMS IN ARITHMETIC. FORM I. GRADE NORMS FOR JUNE TESTING.

	GRADE					
	III	IV	V	VI	VII	VIII
No. of pupils	4181	4589	7142	5927	6632	5269
25-Percentile	3.4	4.2	5.7	5.9	7.6	7.7
Median	3.8	4.6	5.9	6.4	7.8	8.2
75-Percentile	4.3	5.2	6.3	6.8	8.3	8.7

Although the three divisions of the scale were constructed so that it was expected that the scores obtained from the different divisions would be comparable, the grade medians given in Table III clearly indicate that the scores are not comparable. The increase in the median scores from the third grade to the fourth grade is 0.8. The increase from the fourth grade to the fifth grade is 1.3. A similar variation is found in the differences between the subsequent grades. Therefore, the scores obtained by the different divisions of the scale are not comparable. The reason for this is that the pupils taking Division Two or Division Three do not have an opportunity to do the problems of the lower divisions. If they did, a number of them would fail to do all of them correctly. Thus, they would receive a score lower than that which they receive when taking only the higher divisions.

TABLE IV. BUCKINGHAM'S SCALE FOR PROBLEMS IN ARITHMETIC. FORM I
GRADE DISTRIBUTIONS FOR JUNE TESTING

Score	Grade					
	III	IV	V	VI	VII	VIII
9.0					328	699
8.5				6	782	1084
8.0		1	4	11	1349	1290
7.5			14	13	2931	1740
7.0			240	775	58	44
6.5		2	1012	1886		
6.0		6	1540	1504		
5.5	2	21	3663	1569		
5.0	131	1069	106	57		
4.5	490	1474	14	12		
4.0	815	863				
3.5	1305	798				
3.0	967	255				
2.5	298	75				
0	173	25	549	94	1184	412
Total	4181	4589	7142	5927	6632	5269
Median	3.8	4.6	5.9	6.4	7.8	8.2

In Table IV the total distributions are given. Evidently a division of the scale higher or lower than that designed for the grade has been used in a few cases. The distributions are significant in that they show that the divisions of the scale are too difficult for the respective grades. The percent of pupils making zero scores in the third, fifth, seventh, and eighth grades is so large that the scale as now published must be considered unsatisfactory for these grades. This condition could be remedied in the case of Division Two and Division Three by giving the next lower division to the pupils who make zero scores. In the case of Division One, the scale will have to be extended downward by adding less difficult problems.

REFERENCES

First Annual Report, Bureau of Educational Research, University of Illinois, pp. 21-22. These pages contain a very brief suggestion of what was done along the line of this scale that seemed to justify its construction, also a short description of the scale.

Buckingham, B.R. "Notes on the Derivation of Scales in School Subjects, with Special Application to Arithmetic," Fifteenth Yearbook of National Society for the Study of Education, Part I, pp. 23-40.

This presents a report of a series of problems which was given to a number of school children in New York and other cities. The results are given and discussed, especially with reference to locating the problems on a scale. Although this scale is not the one now in use, it is similar to it.

C. MONROE'S DIAGNOSTIC TESTS IN ARITHMETIC.

Monroe's Diagnostic Tests in Arithmetic consist of four parts. Part I includes Tests 1 to 6. Part II includes Tests 7 to 11. These tests involve only integers. Part III includes Tests 12 to 16, which consist of examples involving common fractions. Part IV includes Tests 17 to 21, which consist of examples involving multiplication and division of decimal fractions. Tables V and VI

TABLE V. MONROE'S DIAGNOSTIC TESTS IN ARITHMETIC. GRADE MEDIANS FOR APRIL TESTING. RATE (NUMBER OF EXAMPLES ATTEMPTED)

	GRADE				
	IV	V	VI	VII	VIII
PART I					
(Approximate number of pupils)	900	480	590	600	600
Test 1	7.2	11.6	13.3	12.6	14.0
Test 2	4.1	7.2	9.3	8.6	9.2
Test 3	3.3	5.0	5.8	5.7	7.2
Test 4	2.0	3.2	4.0	4.7	5.7
Test 5	3.9	4.8	5.6	5.7	6.2
Test 6	1.7	2.7	3.1	3.0	4.0
PART II					
(Approximate number of pupils)	380	760	610	520	460
Test 7	3.8	4.2	5.5	5.3	6.3
Test 8	3.0	4.1	5.5	6.1	6.6
Test 9	4.8	5.9	8.2	8.6	9.8
Test 10	2.8	3.2	5.3	5.3	6.7
Test 11	1.6	2.2	2.2	2.9	3.7
PART III					
(Approximate number of pupils)		370	1000	580	560
Test 12		5.8	7.6	8.6	9.4
Test 13		4.5	5.4	6.0	6.0
Test 14		5.2	7.1	8.1	8.7
Test 15		6.2	7.1	7.7	8.1
Test 16		5.7	7.4	8.0	9.1
PART IV					
(Approximate number of pupils)			440	900	660
Test 17			3.6	3.5	4.5
Test 18			11.9	11.5	12.9
Test 19			5.8	4.5	5.3
Test 20			12.5	11.1	13.5
Test 21			5.1	4.3	4.8

give the median scores for these tests in terms of rate (number of examples attempted) and accuracy (percent of examples done correctly). In order to simplify the administration of these tests the plan of scoring has been changed so that the pupil is now given only one score, the number of examples right. In Table VII tentative grade norms are given in terms of this score.

In the interest of economy, both of cost of the tests and time required for their administration, most of the tests of this series were made so short that there is a lack of discrimination between pupils. For example, the increase in the number of examples attempted from grade to grade is frequently less than one example. The shortness of the tests also makes the errors of measurement relatively large.

This group of tests was designed for diagnostic purposes, i. e., it was intended to measure separately the abilities of pupils to do the important types of

TABLE VI. MONROE'S DIAGNOSTIC TESTS IN ARITHMETIC. GRADE MEDIANS FOR APRIL TESTING. ACCURACY (PERCENT OF EXAMPLES CORRECT)

	Grade				
	IV	V	VI	VII	VIII
Part I					
Approximate number of pupils	900	480	590	600	600
Test 1	100	100	100	100	100
Test 2	66.6	86.8	100	100	100
Test 3	56.5	72.3	80.8	82.0	87.6
Test 4	28.0	55.1	71.9	79.8	85.4
Test 5	52.5	61.7	66.9	67.9	75.1
Test 6	22.4	49.5	64.0	77.5	100
Part II					
Approximate number of pupils	380	760	610	520	460
Test 7	63.2	65.1	75.9	76.3	81.6
Test 8	30.4	52.9	66.9	79.8	78.5
Test 9	75.0	86.3	91.2	93.1	100
Test 10	35.2	58.4	72.1	72.3	81.9
Test 11	22.4	35.5	53.4	65.0	68.2
Part III					
Approximate number of pupils		370	1000	580	560
Test 12		35.5	32.0	33.6	49.0
Test 13		38.0	29.2	36.0	53.9
Test 14		57.5	70.3	79.6	86.0
Test 15		37.5	30.0	33.6	45.5
Test 16		38.5	36.8	59.1	70.2
Part IV					
Approximate number of pupils			440	900	660
Test 17			37.6	36.4	53.2
Test 18			100	100	100
Test 19			39.6	47.0	61.7
Test 20			100	100	100
Test 21			35.6	44.0	51.3

examples in the field of the operations of arithmetic. A weighted sum of a pupil's scores on such a group of tests would yield a general measure of his ability in this field.[3]

TABLE VII. MONROE'S DIAGNOSTIC TESTS IN ARITHMETIC. GRADE MEDIANS FOR APRIL TESTING. NUMBER OF EXAMPLES CORRECT

	GRADE				
	IV	V	VI	VII	VIII
PART I					
Approximate number of pupils	900	480	590	600	600
Test 1	7.2	11.6	13.3	12.6	14.0
Test 2	2.8	6.2	9.3	8.6	9.2
Test 3	1.9	3.6	4.7	4.7	6.4
Test 4	.6	1.7	2.9	3.6	4.8
Test 5	2.0	3.0	3.7	3.8	4.4
Test 6	.4	1.3	2.0	2.3	4.0
PART II					
Approximate number of pupils	380	760	610	520	460
Test 7	2.4	2.7	4.2	4.0	5.1
Test 8	.9	2.2	3.7	4.8	5.2
Test 9	3.6	5.1	7.5	8.0	9.8
Test 10	1.0	1.8	3.8	3.8	5.5
Test 11	.4	.8	1.2	1.8	2.4
PART III					
Approximate number of pupils		370	1000	580	560
Test 12		2.0	2.4	2.9	4.6
Test 13		1.7	1.6	2.2	3.2
Test 14		3.0	5.0	6.5	7.4
Test 15		2.3	2.1	2.6	3.7
Test 16		2.2	2.7	4.7	6.4
PART IV					
Approximate number of pupils			440	900	660
Test 17			1.4	1.3	2.4
Test 18			11.9	11.5	12.9
Test 19			2.4	2.1	3.3
Test 20			12.5	11.1	13.5
Test 21			1.8	1.9	2.5

REFERENCES

Finley, G. W., *A Comparative Study of Three Diagnostic Arithmetic Tests.* Colorado State Teachers College Bulletin, Series XX, No. 4.

This reports a study made of the Cleveland Survey Tests, The Woody Arithmetic Scales and Monroe's Diagnostic Tests in Arithmetic. The tests were given on six successive days to some 60 eighth grade pupils. The scores made are given in detail, compared with each other and with scores obtained elsewhere.

Monroe, W. S., "A Series of Diagnostic Tests in Arithmetic," *Elementary School Journal*, XIX, (April, 1919), 585-607.

This article discusses the types of examples in the four fundamental operations, the question of "one dimensional" vs. "two dimensional" tests, and thus establishes the theoretical bases of the tests presented. The series is described, a distribution of scores made thereon is analyzed, and the value of using such tests pointed out.

---

[3] See the group of tests on the operations of arithmetic included in the Illinois Examination.

Uhl, W. L., "The Use of Standardized Materials in Arithmetic for Diagnosing Pupils' Methods of Work," *Elementary School Journal*, XVIII, (November, 1917), 215-218.
This article contains no reference to the Monroe Tests, but describes an experiment in diagnosis similar to that made possible by their use. Both finding specific faults and remedying them is considered briefly.

D. MONROE'S STANDARDIZED SILENT READING TESTS.

Monroe's Standardized Silent Reading Tests have been used so widely that a detailed description here is unnecessary. Each test consists of several exercises, each of which has been assigned a rate value and a comprehension value. The rate value is based upon the number of words in the exercise and the comprehension value is based upon the rate and accuracy with which pupils were found to be able to do the exercise. Test I is for Grades III, IV, and V, Test II is for Grades VI, VII, and VIII, and Test III is for the high school. There are three forms of Tests I and II. There are only two forms of Test III.

The different forms of these tests were constructed so that they were expected to be equivalent. The use of the forms, however, indicates that they are not equivalent. In order to study the degree of equivalence of the three forms, copies of the different forms were arranged in alternate order before the test papers were distributed to the pupils. This plan results in the first, fourth, seventh, tenth, etc. pupil having a copy of Form 1. The second, fifth, eighth, eleventh, etc. pupil would have a copy of Form 2. The third, sixth, ninth, twelfth, etc. pupil would have a copy of Form 3. By this plan each form of the test is given to similar samples of the school population.

Test I was given to approximately 775 pupils and Test II was given to approximately 645. The numbers of pupils taking the different forms in each grade differed slightly. This is an accidental result of the way in which the test papers were arranged. The average and the standard deviation have been calculated for each distribution of scores. In general the pupils made higher scores on Forms 2 and 3 than they did on Form 1. The standard deviations are also unequal. This suggests that the exercises of the different forms of the tests make somewhat irregular scales.

The formula for reducing the scores obtained from one scale to equivalent scores on another scale is as follows:

$$S_1 = \frac{\sigma_1}{\sigma_2} S_2 + \left(Av_1 - \frac{\sigma_1}{\sigma_2} Av_2 \right)$$

In this formula S_1 is the equivalent score in Form 1 and S_2 the obtained sorec in Form 2. Av_1 refers to the average of the scores obtained from Form 1, Av_2 refers to the average of the scores obtained from Form 2. σ_1 is the standard deviation of the distribution of the Form 1 scores and σ_2 is the standard deviation of the distribution of the Form 2 scores. This formula is based upon the usual assumption that the deviations from the average are equal when expressed in terms of the standard deviation of the distribution; in other words that

$$\frac{S_1 - Av_1}{\sigma_1} = \frac{S_2 - Av_2}{\sigma_2}$$

When this equation is solved for S_1 we obtain the formula as given above. Since the scores on Form 1 are in general smaller than the scores on the other two forms it was decided to reduce both the Form 2 and the Form 3 scores to the equivalence of Form 1 scores. The application of the above formula involves the determination of the numerical value of the ratio of $\frac{\sigma_1}{\sigma_2}$ by which the Form 2 score is to be multiplied and the determination of the numerical equivalent of the constant terms of the formula, (i.e., of the expression in parentheses). This latter numerical equivalent may be plus or minus. When it is positive it is to be added and when negative it is to be subtracted.

TABLE VIII. MONROE'S STANDARDIZED SILENT READING TESTS

AVERAGES, STANDARD DEVIATIONS AND CORRECTION NUMBERS TO REDUCE FORM 2 AND FORM 3 TO THE EQUIVALENCE OF FORM 1.

	FORM	NO. PUPILS	AV.	STANDARD DEVIATION	CORRECTION NUMBERS	
					MULTIPLIER	CONSTANT TERM
COMPREHENSION						
TEST I	1	775	11.4	6.5	-	-
	2	777	12.0	5.9	1.100	-1.8
	3	769	12.0	6.2	1.040	-1.1
TEST II	1	649	22.2	9.5	-	-
	2	643	23.8	9.0	.944	-2.6
	3	639	25.2	9.6	.885	- .1
RATE						
TEST I	1	775	63.8	24.0	-	-
	2	777	72.7	27.5	.870	+ 6
	3	769	80.4	29.1	.820	-2.1
TEST II	1	649	97.2	29.5	-	-
	2	643	117.7	32.9	.896	-8.3
	3	639	104.3	22.7	1.300	-38.4

Formula: $S_1 = \frac{\sigma_1}{\sigma_2} S_2 + (Av_1 - \frac{\sigma_1}{\sigma_2} Av_2)$

In the above table the first correction number or multiplier is $\frac{\sigma_1}{\sigma_2}$ and the second or constant term is

$$(Av_1 - \frac{\sigma_1}{\sigma_2} Av_2).$$

In Table VIII, the number of pupils, the average score and the standard deviation is given for each form of each test. In the last two columns the multiplier and the constant term in the above formula are given for Form 2 and Form 3. These can be used in reducing scores obtained from Form 2 or Form 3 to the basis of Form 1. In securing data for these determinations, each test was given in each of the three grades for which it is intended. Except in the eighth grade the number of pupils in the different grades was approximately the same. The correction numbers were calculated for each grade separately. Since they were found to be approximately the same it was decided to combine the scores from the different grades and compute a single set of correction numbers for each test.

The grade medians calculated from the distributions of the scores yielded by the different forms furnished additional information concerning the degree of their equivalence. This information is not in complete agreement with that obtained by the study described. Although it is less accurate it deserves some consideration in the formulation of a set of rules for translating the scores obtained from one form to the basis of another form. In Table IX-A grade medians for all forms are given, and in Table IX-B correction numbers which may be used in reducing scores from Form 2 or Form 3 to Form 1. The correction numbers are based primarily on the results of the study just described but some weight was given to the information furnished by the tabulations of

TABLE IX-A. MONROE'S STANDARDIZED SILENT READING TESTS. GRADE MEDIANS FOR JANUARY AND JUNE TESTING, BASED UPON 130,000 SCORES

	Grade									
	III	IV	V	VI	VII	VIII	IX	X	XI	XII
Form 1										
Rate										
January	52	70	87	90	100	106	83	85	90	96
June	60	79	94	96	104	108	86	87	94	100
Comprehension										
January	6.8	12.7	17.8	18.5	22.8	26.0	23.0	25.4	27.2	30.0
June	9.3	15.3	20.8	21.0	24.5	27.3	24.0	26.0	28.6	32.0
Form 2										
Rate										
January	63	77	98	116	130	133	84	90	98	104
June	70	88	106	124	132	136	86	92	101	109
Comprehension										
January	8.3	13.3	17.2	18.1	26.0	28.2	25.4	28.0	31.0	33.1
June	10.6	15.6	20.5	20.8	27.3	29.4	26.6	29.4	32.2	34.5
Form 3										
Rate										
January	78	92	97	101	109	111				
June	85	95	104	106	111	114				
Comprehension										
January	9.3	14.8	18.4	22.2	26.5	29.8				
June	11.9	16.8	21.5	24.4	28.2	30.5				

TABLE IX-B. APPROXIMATE CORRECTIONS BY WHICH TO MULTIPLY FORM 2 AND FORM 3 SCORES TO REDUCE TO THE BASIS OF FORM 1 SCORES

	Rate			Comprehension		
	Test I	Test II	Test III	Test I	Test II	Test III
Form 2	.88	.80	.94	.95	.93	.90
Form 3	.78	.93		.94	.86	

the scores obtained from the different forms. This is the explanation of some apparent inconsistencies in the reductions to the basis of Form 1.

It should be noted that the scores of the different tests in this series are not comparable. This is to be expected in the case of the rate scores but in the case of the comprehension scores an effort was made to have the different tests yield comparable scores. This attempt was not successful.

The grade distributions which are not published here, show that the tests are too short for the time allowed. In order to secure accurate measures of the abilities of the best readers it will be necessary either to lengthen the test or to shorten the time allowed. The wide spread use of these tests has revealed other defects. Instead of attempting to remedy these defects in the present series it was decided to derive an entirely new series. These have been issued under the title of "Monroe's Standardized Silent Reading Tests, Revised." Three forms of Tests I and II are now available. They were originally published as a part of the Illinois Examination but are now printed separately.

In Tables X-A and X-B we have assembled a miscellaneous collection of grade medians. These are published because a number of requests have been received for just this type of information.

REFERENCES

Monroe, W. S., "Monroe's Standardized Silent Reading Tests," *Journal of Educational Psychology*, IX, (June, 1918), 303-312.

The deriviation of the tests more or less based upon the Kansas Silent Reading Test plan, is briefly sketched. The investigation of weighting and timing is outlined, a sample of the tests is given and some few data concerning results from pupils.

Barnes, Harold, "Reorganization of Classes Based on the Monroe Silent Reading Tests," *University of Pennsylvania Bulletin.* vol. XX, No. 1, 119-123.

This article recounts the use made of these tests in the elementary grades of Girard College. Not only are the scores presented, but also the resulting organization upon the basis of ability as shown on the tests is described.

Kelly, F. J., "Kansas Silent Reading Tests," *Journal of Educational Psychology*, VII, (February, 1916), 63-80.

The author of these tests, which were the forerunners of Monroe's Standardized Silent Reading Tests, gives a brief statement of the construction, administration, and use of the tests, following it with a more detailed statement of results secured in nineteen Kansas cities.

Lloyd, S. M. and Gray, C. T., "Reading in a Texas City, Diagnosis and Remedy," *University of Texas Bulletin*, No. 1853.

This bulletin gives an account of a study of the reading situation in Austin. The Monroe tests were given in grades 3-7. Results obtained are analyzed at considerable length, measures to improve the situation are discussed, and improvement after a period of special emphasis on reading is shown.

Pressey, S. L. and L. W., "The Relative Value of Rate and Comprehension Scores in Monroe's Silent Reading Test, as Measures of Reading Ability," *School and Society*, (June 19, 1920), 747-49.

In a brief discussion of the above subject, the writers present results of correlating teachers' estimates of reading ability with rate and comprehension scores, also the latter with each other. They conclude that comprehension scores may tell us all the tests can about children's ability in reading.

TABLE X-A. MONROE'S STANDARDIZED SILENT READING TESTS
GRADE MEDIANS, RATE

	GRADE									
	III	IV	V	VI	VII	VIII	IX	X	XI	XII
FORM 1										
Large Cities (May)	57.0	70.5	88.1	91.6	110.0	106.2				
Medium Sized Cities (May)	43.6	65.0	78.5	88.7	94.0	106.4	82.5	85.6	86.6	102.6
FORM 2										
Large Cities (May)	64.1	77.0	97.1	114.9	133.7	138.3				
Medium Sized Cities (May)	60.1	74.5	97.9	113.2	130.1	135.9	87.4	88.0	89.2	101.6
Wisconsin Cities (February)	69.0	76.0	103.0	115.0	133.0	133.0	89.0	89.0	106.0	106.0
Denver (October)	31.0	59.0	76.0	81.0	98.0	98.0				
Denver (May)	53.7	76.0	95.0	115.0	127.5	133.2				
Pittsburgh (May)	65.0	78.0	93.0	105.0	117.0	133.0				
FORM 3										
Large Cities (May)	81.4	101.6	111.0	102.7	113.5	119.0				
Medium Sized Cities (May)	80.1	93.3	105.4	100.9	108.8	112.8				
FORM NOT SPECIFIED										
Pennsylvania Cities and Boroughs (April)*	63.0	78.0	89.0	105.0	115.0	136.0				

*See *Seventh Annual Schoolmen's Week Proceedings*, University of Pennsylvania Bulletin, Vol. XXI, No. 1, p. 122.

TABLE X-B. MONROE'S STANDARDIZED SILENT READING TESTS GRADE MEDIANS, COMPREHENSION

	GRADE									
	III	IV	V	VI	VII	VIII	IX	X	XI	XII
Form 1										
Large Cities (May)	7.9	12.1	18.9	18.6	24.6	27.3				
Medium Sized Cities (May)	4.9	12.0	14.6	18.8	20.6	24.1	24.0	25.8	28.5	33.3
Form 2										
Large Cities (May)	9.2	13.7	17.8	23.5	27.7	30.3				
Medium Sized Cities (May)	7.7	13.1	17.9	21.7	25.3	29.3	23.8	27.2	29.6	31.6
Wisconsin Cities (February)	8.9	13.4	18.6	21.4	25.0	27.9	26.5	29.8	32.9	37.5
Denver (October)	2.0	7.5	11.0	16.0	19.0	21.0				
Denver (May)	8.0	12.5	16.0	22.0	24.5	30.0				
Pittsburgh (May)	8.5	13.5	16.0	18.5	23.5	28.5				
Form 3										
Large Cities (May)	10.9	14.5	19.5	23.2	26.6	32.5				
Medium Sized Cities (May)	9.6	15.2	19.0	22.4	26.5	31.0				
Form not specified										
Pennsylvania Cities and Boroughs (April)*	8.3	12.1	15.7	18.0	22.9	27.6				

*See *Seventh Annual Schoolmen's Week Proceedings*, Univeristy of Pennsylvania Bulletin Vol. XXI, No. 1, p. 122.

E. CHARTERS DIAGNOSTIC LANGUAGE TESTS, AND DIAGNOSTIC LANGUAGE AND GRAMMAR TESTS.

There are two groups of these tests: (1) The Diagnostic Language Tests, designed for Grades III to VIII inclusive, which include, Pronouns, Verbs, (formerly Verbs A), Miscellaneous A (formerly Miscellaneous), Miscellaneous B (formerly Verbs B); (2) The Language and Grammar Tests, designed for Grades VII and VIII, which include Pronouns, Verbs (formerly Verbs A) and Miscellaneous A (formerly Miscellaneous). The Language Tests consist of a number of sentences most of which are grammatically incorrect. If a sentence is correct the pupil makes a cross on the dotted line below the sentence. If the sentence is not right the pupil is required to put the correct words on the dotted line below it. In the Language and Grammar Tests the pupil is required in addition to write the rule on which the correction is based. The pupil's score is the number of exercises which he does correctly. Since the sentences which make up the tests were selected as representative of the errors which pupils make, a pupil's performance on the tests gives a diagnosis of his abilities in the field of these tests.

There are two forms of these tests. The second form, however, was not published until September, 1920. Consequently, the scores reported in this bulletin are based on Form 1. Although the two forms were constructed so that Form 2 might be expected to be equivalent to Form 1, there is available at this time no information concerning the degree of their equivalence.

TABLE XI A. GRADE NORMS FOR CHARTERS' DIAGNOSTIC LANGUAGE TESTS. MARCH TESTING

	GRADES					
	III	IV	V	VI	VII	VIII
*MISCELLANEOUS A						
Number of Pupils	386	669	668	845	758	494
25-percentile	4.0	5.8	8.1	11.8	14.0	16.6
Median	6.7	9.3	11.6	16.5	18.9	22.3
75-percentile	13.3	13.6	16.0	21.7	24.4	27.1
†MISCELLANEOUS B						
Number of Pupils	230	430	307	475	412	294
25-percentile	3.0	10.6	15.7	19.8	23.5	28.7
Median	7.9	17.8	22.0	27.3	29.4	32.0
75-percentile	14.8	24.5	27.6	32.4	33.7	36.8
**VERBS						
Number of pupils	365	403	373	478	539	638
25-percentile	7.3	12.9	17.2	19.0	22.7	28.6
Median	12.6	17.7	22.6	24.3	27.7	32.8
75-percentile	18.8	22.7	28.4	29.3	31.9	36.1
PRONOUNS						
Number of pupils	787	864	895	1344	1566	1253
25-percentile	8.9	11.1	14.2	17.0	19.6	23.1
Median	13.6	15.1	18.5	21.4	24.5	29.0
75-percentile	19.8	20.3	22.6	25.7	29.5	34.0

* Formerly Miscellaneous
† Formerly Verbs B
** Formerly Verbs A

TABLE XI-B. GRADE NORMS FOR CHARTERS' DIAGNOSTIC LANGUAGE AND GRAMMAR TESTS MARCH TESTING

	GRADES	
	VII	VIII
MISCELLANEOUS		
Number of pupils	332	362
25-percentile	2.9	6.1
Median	6.3	11.9
75-percentile	11.7	18.7
VERBS		
Number of pupils	434	497
25-percentile	2.8	6.9
Median	7.8	14.0
75-percentile	22.9	24.1
PRONOUNS		
Number of pupils	332	362
25-percentile	4.4	8.5
Median	8.0	17.1
75-percentile	16.7	26.0

REFERENCES

Charters, W. W., "Minimum Essentials in Elementary Language and Grammar," *Sixteenth Yearbook of the National Society for the Study of Education.* Part I, 85-110.

This article gives a brief account of a number of studies of language and grammar errors made by school children, with tables of results. These studies were the basis of the content of Charters' tests.

Sixth Conference on Educational Measurements. Bulletin of the Extension Division, Indiana University, Vol. V, No. I, pp. 6-12 and 13-24.

These two discussions by Charters give a rather general discussion leading up to a brief account of the development and form of the tests, followed by some suggestions as to their use.

Charters, W. W., "Constructing a Language and Grammar Scale," *Journal of Educational Research*, I (April, 1920), 249-257.

The tests herein considered are a revision of those referred to above. The writer gives a short description of their derivation, use, scoring, etc. The question of weighting is discussed and the reason for its elimination given.

F. WILLING'S SCALE FOR MEASURING WRITTEN COMPOSITION

The Willing Scale for Measuring Written Composition differs from other composition scales in that an attempt is made to secure separate measures of "form value" and "story value." The "form value" of a pupil's composition is based upon his errors in grammar, punctuation, capitalization, and spelling. In order to make the scores in form value comparable, the number of errors which the pupil makes is multiplied by 100 and divided by the number of words in his composition. The quotient is the number of errors per hundred words. The "story value" of a pupil's composition is its value when errors of grammar punctuation, capitalization, and spelling are neglected. This value is measured by means of the scale.

TABLE XII. GRADE NORMS FOR WILLING'S SCALE FOR MEASURING WRITTEN COMPOSITION, MARCH TESTING

	Grade					
	III	IV	V	VI	VII	VIII
Approx. No. of Pupils	325	580	705	695	570	130
Story Value						
25-percentile	30.5	43.4	55.7	60.9	65.9	65.3
Median	41.5	58.7	68.1	74.0	76.6	79.0
75-percentile	54.8	74.5	78.8	85.2	86.6	86.2
Form Value (Errors per 100 Words)						
25-percentile	11.7	6.2	3.6	3.2	2.6	2.3
Median	18.5	10.7	6.8	5.8	4.4	4.4
75-percentile	26.0	17.3	10.9	9.8	6.4	7.0

The grade distributions given in Tables XIII and XIV indicate that the scale needs to be extended at both ends. It does not contain steps low enough in story value to provide adequate measures for many compositions contributed by pupils in the third and fourth grades. Neither does it provide adequate measures for the best compositions in grades beyond the fourth. For practical purposes these limitations are not serious because when a pupil's composition is as poor as 20 on this scale the pupil needs special attention. When a pupil writes a composition as good as 90 on this scale it is likely that special instruction is superfluous. In addition it may be pointed out that the median score of a class is probably not affected by these limitations of the scale.

TABLE XIII. WILLING'S SCALE FOR MEASURING WRITTEN COMPOSITION. GRADE DISTRIBUTIONS FOR MARCH TESTING

Story Value Score*	Grade					
	IV	V	VI	VII	VIII	IX
90	7	28	52	81	95	14
80	6	67	105	177	141	48
70	8	91	168	147	150	25
60	28	91	146	123	107	21
50	60	98	110	75	41	12
40	60	90	77	49	31	4
30	75	66	39	32	5	4
20	76	48	8	8	2	1
Total	320	579	705	692	572	129
Median	41.5	58.7	68.1	74.0	76.6	79.0

*These intervals are expressed in terms of their midpoints.

References

Wllling, M. H., "The Measurement of Written Composition in Grades IV to VIII. *English Journal*, VII (March, 1915), 193-202.

The writer explains and outlines the measurement of written composition especially in connection with the Denver and Grand Rapids surveys. The method of constructing the scale, its use, scoring, results obtained, etc. are discussed, and the scale reproduced.

The *Denver Survey*, 1916, Part II, pp. 59-63, and the *Grand Rapids Survey*, 1916, pp. 85-105, give accounts of the use of this scale. The latter contains rather complete tables and graphs of pupil achievement, and comparisons of results with those obtained in Denver

TABLE XIV. WILLING'S SCALE FOR MEASURING WRITTEN COMPOSITION. GRADE DISTRIBUTIONS FOR MARCH TESTING

Form Value*	GRADE					
	IV	V	VI	VII	VIII	IX
30	50	14	3	5	2	
27	24	17	1	4	3	
24	22	27	8	4	2	
21	32	29	9	4	1	
18	43	38	19	18	4	
15	29	51	45	17	15	2
12	42	56	52	48	28	4
9	34	120	105	96	49	12
6	32	79	149	140	131	21
3	14	98	170	199	200	49
0	5	41	141	158	138	41
Total	327	570	702	693	573	129
Median	18.5	10.7	6.8	5.8	4.5	4.4

* Errors per 100 words.

G. HARLAN'S TEST FOR INFORMATION IN AMERICAN HISTORY

This test consists of ten exercises in the field of American History and is designed for use in the seventh and eighth grades. Each exercise consists of two or more parts. The maximum score which a pupil may receive is 100.

TABLE XV. GRADE NORMS FOR HARLAN'S TEST OF INFORMATION IN AMERICAN HISTORY. MAY TESTING

	GRADE	
	VII	VIII
Number of pupils	1109	1691
25-percentile	30.1	45.7
Median	43.9	68.2
75-percentile	57.3	83.3

In Table XVI the distributions of scores for the seventh and eighth grades are given. These distributions are of interest because of the very great individual differences which they suggest. It is possible that the apparent differences are due in a considerable measure to the errors of measurement. Since there is only one form of the test no measure of reliability is available.

TABLE XVI. HARLAN'S TEST OF INFORMATION IN AMERICAN HISTORY. GARDE DISTRIBUTIONS FOR MAY TESTING

Score	Grade	
	VII	VIII
96	6	66
91	3	118
86	13	140
81	18	187
76	35	136
71	35	136
66	54	112
61	65	92
56	65	100
51	99	95
46	113	79
41	115	98
36	100	90
31	93	79
26	94	61
21	79	42
16	63	35
11	37	13
6	19	10
0	3	2
Total	1109	1691
Median	43.9	68.2

References

Harlan, Chas. L., "Educational Measurement in The Field of History," *Journal of Educational Research,* II (December, 1920), 849-853.

The writer follows a short discussion of tests in the "content" subjects with a brief description of his test and its use in nine cities. The requirements he deems essential to a good test are listed as a basis of his test.

Griffith, G. L., "Harlan's American History Tests in the New Trier Township Schools," *School Review* (November, 1920), 697-708.

The first half of this article is devoted to a general discussion of history and history testing. This is followed by a description of the test and the results of its use in the eighth grade of this township. Data are given for each of the single exercises of the test.

H. SACKETT'S SCALE IN UNITED STATES HISTORY

This scale, arranged by L. W. Sackett, was originally devised by Bell and McCollum. It consists of seven tests which appear to have been intended for use in secondary schools and colleges. The medians given in Table XVII are for the eighth grade. The number of scores is such that it is doubtful if the median scores have much value for use as tentative standards.

TABLE XVII. GRADE NORMS FOR SACKETT'S SCALE IN UNITED STATES HISTORY. MAY TESTING. EIGHTH GRADE

	Tests						
	I	II	III	IV	V	VI	VII
Number of pupils	111	101	107	92	93	78	85
25-percentile	62.6	50.8	69.2	37.0	44.7	7.5	9.6
Median	118.7	146.2	115.0	125.0	86.5	46.6	96.5
75-percentile	192.5	273.9	183.1	287.5	193.7	138.1	195.5

References

Bell, J. C. and McCollum, D. F. "A Study of the Attainments of Pupils in United States History," *Journal of Educational Psychology*, VIII (May, 1917), 257-74.

The writers follow a discussion of historical ability with an account of the use of test material in various schools from Grade V through the senior year of the University of Texas. The results secured are analyzed. The test questions used were in general similar in kind to those of Sackett's Scale in Ancient History, although based upon United States History.

Sackett, L. W. "A Scale in Ancient History." *Journal of Educational Psychology*, VIII (May, 1917), 284-93.

The test questions are given with a brief statement of their source, use, and scoring. Results are given from almost 1000 papers, and the relative difficulty of the questions computed.

Sackett, L. W. "A Scale in United States History," *Journal of Educational Psychology*, X (September, 1919), 345-348.

The writer tells of the development of this scale out of the data furnished by Bell and McCollum's work referred to above. The determination of the relative difficulty of the parts is given considerable space.

I. Hotz's First Year Algebra Scale

This scale consists of five separate scales: (1) Addition and subtraction; (2) Multiplication and division; (3) Equation and formulae; (4) Problems; (5) Graphs. Each sub-scale consists of exercises arranged in order of increasing difficulty.

References

Hotz, H. G. *First Year Algebra Scales*, Teachers College, Columbia University, Contributions to Education No. 90.

The writer gives a history of the derivation of these scales, a complete reproduction of them, and a discussion of their administration and use. The statistical working out of the scales is treated fully for one of them, the procedure for all being the same.

Cawl, F. R. "Practical Uses of an Algebra Standard Scale," *School and Society* (July, 1919), 89-91.

The results of testing a class in a large private school are here presented. The matter of correlation with English, French, and Latin is considered. A short interpretation of results is given, with suggestions as to the value of using such a scale.

TABLE XVIII. GRADE NORMS FOR HOTZ'S FIRST YEAR ALGEBRA SCALES. MAY TESTING

	Grade	
	IX	X
ADDITION AND SUBTRACTION		
Number of pupils	561	390
25-percentile	5.2	5.8
Median	6.9	7.3
75-percentile	9.1	8.7
MULTIPLICATION AND DIVISION		
Number of Pupils	570	388
25-percentile	5.7	5.9
Median	7.2	7.4
75-percentile	8.4	8.7
EQUATIONS AND FORMULAS		
Number of Pupils	478	385
25-percentile	6.2	6.7
Median	7.7	7.9
75-percentile	9.7	9.1
PROBLEMS		
Number of Pupils	566	394
25-percentile	4.5	3.9
Median	6.4	5.0
75-percentile	8.6	6.3
GRAPHS		
Number of Pupils	121	413
25-percentile	5.2	4.1
Median	6.2	5.0
75-percentile	7.0	6.0

J. MINNICK'S GEOMETRY TESTS

This series of tests is based on the assumption that the demonstration of a geometrical theorem involves the following abilities: Test A, the ability to draw the figure. Test B, the ability to state the hypothesis and conclusion. Test C, the ability to recall the facts concerning the figure. Test D, the ability to select and organize facts so as to produce the proof. Test E, the ability to draw auxiliary lines. The series includes one test for each of these abilities. No report is made for Test E. These tests are unique in that they provide for both positive scores and negative scores. The positive score is the percent of the necessary elements of the proof given correctly by the pupil. The negative score is the number of incorrect and unnecessary elements.

REFERENCES

Minnick, J. H. *An Investigation of Certain Abilities Fundamental to the Study of Geometry.* University of Pennsylvania.

This monograph gives a synopsis of methods and results used in deriving the tests, followed by a more detailed statement. The latter includes a reproduction of the tests, tables giving data secured from testing, statistical methods of weighting exericises, suggestions as to use, etc.

TABLE XIX. GRADE NORMS FOR MINNICK'S GEOMETRY TESTS

	Positive Scores Grade		Negative Scores Grade	
	X	XI	X	XI
TEST A (ability to draw accurate figures for theorems)				
Number of pupils	126	66	126	60
25-percentile	53.3	43.8	2.4	1.1
Median	63.0	58.0	4.1	2.6
75-percentile	67.2	69.2	6.6	5.4
TEST B (Ability to state hypothesis and conclusion in terms of given figure.)				
Number of Pupils	167	66	167	66
25-percentile	55.2	55.5	1.1	1.0
Median	69.6	67.1	2.3	2.0
75-percentile	81.3	83.6	3.9	3.9
TEST C (Ability to recall known facts about figures when one or more are given).				
Number of Pupils	154	65	154	63
25-percentile	52.2	55.6	1.9	1.4
Median	64.1	64.7	3.8	3.9
75-percentile	77.2	77.9	7.1	5.7
TEST D (Ability to organize and select facts to produce a proof).				
Number of Pupils	155	68	155	54
25-percentile	68.0	75.0	.8	.8
Median	85.5	89.2	1.6	1.6
75-percentile	92.9	98.3	2.3	3.2

Minnick, J. H. "A Scale for Measuring Pupil's Ability to Demonstrate Geometrical Theorems," *School Review*, (Feb., 1919), 101-109.

A brief account of the construction of a scale to measure one definite geometric ability is given. The scores made upon the first selection of exercises, the resultant weighting and then the selection of those best suited to make up a scale are briefly treated. The exercises chosen are reproduced.

Minnick, J. H. "Certain Abilities Fundamental to the Study of Geometry," *Journal of Educational Psychology*, (Feb., 1918), 83-90.

Four abilities requisite to formal geometrical demonstration are listed. Their relation to teaching, development by teaching, and diagnosis by tests are discussed. The tests used were those of the author. Correlations with teachers marks are given.

K. HOLLEY'S SENTENCE VOCABULARY SCALE

This scale consists of a number of exercises of the following type:

1. *Impolite* people are........kindly........brave........young........ill-bred.
2. A man is *afloat* in a........mine..........tower........boat..........hospital.

The pupil is asked to underline the word which makes the truest sentence. These exercises are arranged in order of increasing difficulty, and a pupil's score is found by subtracting one-third of the number of errors from the number correct. An abbreviated form of this scale has been incorporated in the Illinois General Intelligence Scale. The scale was constructed to provide a suitable means of ascertaining the general intelligence of groups of children. The measure which it yields is not sufficiently accurate to be used as an index of the general intelligence of individual pupils. The scale is also recommended as an instrument for measuring the vocabulary of pupils. The total distributions given in Table XXI indicate that this scale is too difficult for pupils in the third and fourth grades.

TABLE XX. GRADE NORMS FOR HOLLEY'S SENTENCE VOCABULARY SCALE APRIL TESTING

	Grade									
	III	IV	V	VI	VII	VIII	IX	X	XI	XII
Number of pupils	406	520	465	450	1188	1047	253	223	155	108
25-percentile	8.4	16.7	25.3	33.4	32.3	40.1	40.9	50.1	52.4	54.5
Median	16.6	25.1	33.0	42.8	41.9	47.7	49.0	56.0	59.9	62.7
75-percentile	28.5	33.6	39.9	51.5	49.7	55.8	57.1	63.5	67.9	70.1

References

Terman, L. M. and Childs, H. G. "A Tentative Revision and Extension of the Binet-Simon Measuring Scale of Intelligence," *Journal of Educational Psychology*, (April, 1912), 205-208.

The basis of Holley's Sentence Vocabulary Scales is the Stanford Revision, 100 word Vocabulary Test, the construction of which is here described. Tentative standards of achievement are also given.

Holley, C. E. *Mental Tests for School Use.* Bureau of Educational Research, University of Illinois, Bulletin No. 4 pp. 86-91.

This bulletin gives an account of a comparative study of six group intelligence scales, of which the above was one, based on data from the school system of Champaign, Illinois. A brief account of the origin of the Sentence Vocabulary Scales is included (p. 30).

Branson, E. P. "An Experiment in Arranging High-School Sections on the Basis of General Ability," *Journal of Educational Research*, (Jan., 1921), 53-56.

At Long Beach, California, this scale was given to two groups of high-school entrants who had recently taken, and been grouped by the Otis Group Intelligence Scale. At the end of the term the test was repeated. A comparison by groups of the scores at the two periods, and correlations with the Otis Scale, are given.

TABLE XXI. HOLLEY'S SENTENCE VOCABULARY SCALE. GRADE DISTRIBUTIONS FOR APRIL TESTING

Score	Grade									
	III	IV	V	VI	VII	VIII	IX	X	XI	XII
90					6					
80					5			1	2	7
70					3	1	8	21	27	21
60		1	5	50	75	126	32	56	48	36
50	31	4	20	74	199	321	79	85	52	31
40	18	44	89	141	379	341	78	45	23	13
30	42	128	170	110	290	153	50	15	3	
20	71	170	137	59	123	73	5			
10	124	130	30	13	74	29	1			
0	120	43	14	3	34	3				
Total	406	520	465	450	1188	1047	253	223	155	108
Median	16.7	25.1	33.0	42.8	41.9	47.8	49.0	56.0	59.9	62.7

L. HOLLEY'S PICTURE COMPLETION TEST FOR PRIMARY GRADES

This test, as it name suggests, consists of a number of pictures which are incomplete. The pupil is expected to add the part which is missing. It was designed as an instrument for measuring the general intelligence of young children. The total distributions as given in Table XXIII indicate that it is not a good instrument for this purpose. The distributions exhibit unusually

TABLE XXII. GRADE NORMS FOR HOLLEY'S PICTURE COMPLETION TEST FOR PRIMARY GRADES. JANUARY TESTING

	Grade				
	Kindergarten	I	II	III	IV
Number of pupils	75	1438	1233	327	167
25-percentile	1.8	4.4	7.9	9.9	8.8
Median	5.3	7.8	11.5	13.5	12.3
75-percentile	7.8	12.2	15.1	16.5	15.0

TABLE XXIII. HOLLEY'S PICTURE COMPLETION TEST FOR PRIMARY GRADES. GRADE DISTRIBUTIONS FOR JANUARY TESTING

Score	Grade			
	I	II	III	IV
20	34	31	9	4
19	23	43	13	4
18	33	46	25	9
17	36	59	20	12
16	42	64	27	3
15	52	72	34	10
14	47	76	22	15
13	54	88	26	14
12	48	98	24	17
11	82	77	19	14
10	72	92	24	9
9	81	84	24	12
8	91	89	26	9
7	103	75	11	11
6	122	59	9	13
5	94	50	8	8
4	112	50	4	2
3	108	29	1	1
2	86	21	1	
1	77	28		
0	41	2		
TOTAL	1438	1233	327	167
MEDIAN	7.8	11.5	13.5	12.3

high variability. This is much greater than is exhibited by other tests when applied to children in these grades. The median scores given in Table XXII give further indications of the inadequacy of this test, particularly in the grades above the first.

REFERENCES

Holley, C. E. *Mental Tests for School Use.* Bureau of Educational Research, University of Illinois, Bulletin No. 4, pp. 86-91.

A general discussion of tests of this type is followed by an account of the testing from which this test came. This was done in Champaign, Illinois. Results are merely outlined.

CHAPTER III

THE DERIVATION OF MONROE'S STANDARDIZED REASONING TESTS IN ARITHMATIC[4]

The process of problem solving. "Reasoning" as it occurs in the solving of an arithmetical problem involves these steps: (1) A careful reading of the problem including the association of correct arithmetical meanings with the "technical" terms used in stating the problem. (2) Recall of facts and principles suggested by the problem and required for its solution. (3) Formulation of a hypothesis or plan of solution using as data the results of the first two steps. (4) Verification of this plan of solution. This process of reasoning is usually followed by the calculations outlined in the plan of solution. This additional step, however, is not a part of the reasoning process.

Two kinds of words are used in stating arithmetical problems: (1) The descriptive words give the setting of the problem. Only in an indirect way do these affect the solution. (2) The "technical terms" of an arithmetical problem consist of those words and phrases which define quantities and quantitative relationships. Every problem involves at least three quantities, two given and the third to be found. These quantities are related in a definite way. For example, the sum of the two quantities given equals the third, or the third is the quotient of one divided by the other. In problems involving two or more steps there are more than three quantities and the relationships are more complex. However, in every case there are words or phrases which either directly or indirectly tell what these relationships are, and, consequently, what operations must be performed to obtain the desired answer.

This principle may be illustrated by the following problems: "What are the average daily earnings of a boy who receives $0.88, $0.25, $1.15, $0.75, $0.50, and $0.60 in one week?"

The phrase "average daily earnings" names the quantity to be found and also specifies its relationship with the given quantities. The "average" is the quotient of the sum of the several amounts divided by the number of items. A knowledge of this definite meaning of "average" is necessary if one is formulating a rational plan of solving the problem. If the phrase "average daily" was omitted we would have an entirely different problem.

"How many square yards of linoleum will be required to cover a floor 16 feet by 12 feet?"

"How many square yards" names the third quantity in this problem and in connection with "15 feet by 12 feet" specifies the relations which exist between the quantities. This third quantity is the product of the dimensions divided by nine.[5] In this case the number of square feet in a square yard must be remembered and also the principle that the area of a rectangle (i.e., a figure whose dimensions are given as in the problem) is the product of the length by the width.

---

[4] A number of considerations on which the derivation of these tests is based are contained in an article by the writer in *School and Society*, Volume VIII, pages 295 and 424. Sample copies of these tests may be obtained from the Public School Publishing Company, Bloomington, Illinois.

In many cases when the first two steps of the reasoning process have been completed satisfactorily, the formulation of the plan of solution (the next step in the reasoning process) involves little uncertainty. In fact it is essentially mechanical. This is the case in these illustrations. In the case of very simple problems, or very familiar problems, the reasoning process is usually short-circuited so that there is no explicit association of meaning with the technical terms nor recall of principles. The problem as a whole or some feature of it serves as a cue for the direct association of the plan of solution. In such cases there is strictly speaking, no reflective thinking or reasoning, and the mental process involved is much the same as that which occurs in the operations of arithmetic. The solution of the problem has become automatic.

The nature of a reasoning test in arithmetic. A reasoning test in arithmetic is essentially a test of careful reading in a limited field to answer specific questions. In this reading, technical vocabulary is fundamental. The pupil gives evidence of his degree of comprehension by his plan of solution. The correctness of the numerical answer to the problem depends upon the accuracy of the pupil's calculations and the recall of denominate number facts as well as upon the plan of solution. The *plan*, or principle, of the solution and not the *accuracy* of the numerical answer is, therefore, the measure of the pupil's ability to reason in arithmetic. Thus in describing a pupil's performance on a reasoning test, errors in the recall of facts and in calculation should be disregarded. For the problems which are solved correctly in principle a score based on correct answers may be used as a crude measure of the pupil's ability to perform the operations of arithmetic.

In order that a pupil's score on a reasoning test may be indicative of his ability to solve arithmetical problems in general, the problems must be carefully selected with reference to content (vocabulary). The ideal reasoning test would be one that included all of the technical terms but this is not possible because the vocabulary of arithmetical problems is extremely varied and voluminous. In another[6] place the writer has reproduced 28 different forms of statement which were found in the examination of eight text-books for the problem, "Given, $7.50, paid for silk, and price per yard $1.50, to find the number of yards purchased." This condition makes it necessary to select a few problems which will be representative in respect to content, in order to have a test of usable length.

---

[5] An alternative solution is to reduce each dimension to yards before finding the area.

[6] Monroe, Walter S. *Measuring the Results of Teaching*, Houghton Mifflin Company (1918), 163.

Method of selecting problems on basis of content. In the case of the series of tests described in this report the representative problems were selected by the following method. The one- and two-step problems appearing in eight widely used texts were classified according to the operation or operations they called for. This gave in one group all the problems requiring only addition, in another those requiring only subtraction and so on. The problems in each of these groups were further classified by the writer on the basis of the technical terms used. This was found to be difficult because of the great variety of these terms. Since the classification represents the judgment of only one person, it cannot be considered final in any sense.

The general plan of classification may be illustrated by the types of division problems. In the list on the following pages only those types are given which included problems found in five or more of the eight texts examined. A large number of additional types included problems from less than a majority of the texts. A descriptive statement of the type is followed by a limited number of illustrative problems. It will be noted that for a single type these problems are not identical in vocabulary but it was the judgment of the writer that they were sufficiently similar to justify grouping them together. It has been assumed that the terms used are essentially synonymous. This hypothesis is, of course, subject to experimental verification. Unfortunately, this is lacking at this time. However, the resulting list of type problems is more representative of the vocabulary of arithmetical problems as they occur in our texts than any other available list.

DESCRIPTION OF TYPES AND ILLUSTRATIVE PROBLEMS:

1. Given a whole to find number of parts of a given size, including to find the number of acres to produce a given yield.

   A baker used three-fifths lbs. of flour to a loaf of bread. How many loaves could he make from a barrel?

   When the average yield per acre is 25 bushels how many acres will yield 925 bushels.

   How many lengths three-fourths yds. long can be cut from 15 yds. of goods?

   How many hens can be properly accommodated in a pen containing 51 square feet, if each hen requires 6 square feet?

   964 marbles are distributed equally among a certain number of boys. Each boy has 82 marbles. How many boys are there?

   At 7½ gallons to the cubic foot, how many cubic feet will 3000 gallons of oil occupy?

   Oats weigh 32 lbs. to a bushel. How many bushels are there in a load weighing 1344 lbs?

2. Given cost and price to find the number of articles purchased. This includes wages when question is how many days, weeks, etc. to earn a given amount.

   At 16 cents per pound, how many pounds of steak does a woman get if the amount of the purchase is 80 cents.

3. The reverse of No. 1. Given whole and number of parts to find size of each part.

Three boys buy a rowboat for twelve dollars and seventy-five cents, sharing the expense equally. Find how much each boy has to pay.

If 54 marbles are divided equally among 6 boys, how many marbles will each receive?

In 28 days a hotel used 361 lbs. of butter. How many pounds did it use a day?

4. The reverse of No. 2.

A farmer paid thirty-three dollars and a half for 4 bushels of seed wheat. How much did he pay for a bushel?

The bill for 58 tons of copper amounted to 612 dollars. What was the price per ton?

A fowl weighing 4 and one half lbs. sells for $1.00. What is the price per pound?

A man's wages amounted to 46 dollars for 9 and one-fifth day's work. How much did he receive per day?

A man works 8 hours a day for 4 dollars and 80 cents. How much does he receive for each hour's work?

5. Given the price for a given denomination to find the price at a lower denomination.

A boy bought a dozen oranges at the rate of 15 cents a dozen. What did they cost him apiece?

When milk is 10 cents a quart, how much is a pint worth?

6. Given the whole and the number of parts to find the average (rate, price, yield, etc.)

A farmer raised 500 bushels of wheat on a field of 40 acres. What was the average yield per acre?

A fast train runs from Chicago to a station 356.4 mi. distant in exactly 9 hours. What is the average rate of the train?

A drover paid $1125 for cows, what was the average price if he bought 25?

A mill employs 600 hands and has a weekly pay roll of $2,000. What is the average weekly wage for each employee?

7. The whole and the rate are given. The question is asked by, "How long?"

If a horse eats three-eights bu. of oats a day, how long will 6 bus. last?

How long will it take to earn 28 dollars at $1.75 a day?

8. Given distance and rate to find how long.

At 25 miles an hour, how long will it take an automobile to go 160 miles?

9. Given distance and number of units of time to find rate.

In 3.2 hours a man walks 12.32 mi. How far does he walk in one hour?

Find the rate of speed per hour made by an airship traveling 218.05 miles in 3.5 hour.

10. A fractional part of a whole is given to find the whole.

If, when 18 and three eighths mi. of track are laid, one third of the road is completed, how long is the road?

I sold a bicycle for 18 dollars. This was three sevenths of what I paid for it. How much did I pay for it?

11. A percent of the whole is given to find the whole.

If 33 and one third percent of a man's loss is 300 dollars, how much does he lose?

A girl spent 25 cents which was 12½ percent of her monthly allowance, how much was her allowance?

A clerk had his weekly wages increased 3 dollars, or 16 and two thirds percent. What were his wages before this increase?

12. Given the amount of gain or profit and percent of gain or profit to find the cost or selling price.

A hardware merchant makes a profit of 25 percent or 32 cents on saws. Find the cost.

A farmer sold his horse at a gain of 30 dollars, or 25 percent. Find the cost.

13. Given the commission and percent of commission to find amount sold.

Five percent commission on a certain amount of money was 684.20 dollars. What was the amount?

14. Given two numbers to find what percent one is of the other.

If 1000 lbs. of potatoes contain 180 lbs. of starch, what percent of potatoes is starch?

If a man saves 187.50 dollars out of his salary of 1250 dollars, what percent does he save?

The boys in the Marshall school won 5 of the 8 games of hockey. What percent?

In his examination in arithmetic a boy had 10 problems out of twelve right. His grade was what percent?

15. Given two numbers to find what part one is to the other.

The Jackson basket-ball team won 35 out of 56 games. What part did it win?

A man spends for rent 360 dollars out of an income of 1500 dollars. What part of his income is spent this way?

16. Given the amount of investment, or principle, and the income or interest to find rate.

Mrs. Lynch received 24 dollars a year interest on 400 dollars loaned Mrs. Burnet. What is the rate?

17. Given an amount in one denomination to reduce to a higher.

An aviator reaches a height of 11,474 feet. Express this height in miles.

A milk dealer sells 302 qts. of cream. Express this as gallons and quarts.

In digging out a cellar 8260 cubic feet of earth were removed. At 27 cubic feet to the cubic yard, how many cubic yds. were removed?

18. Given the value or face of a policy and premium to find the rate.

Find the rate, given the face of the policy as 1500 dollars and premium 15 dollars.

A fire insurance company charged 20 dollars for insuring an automobile for 1000 dollars. What was the rate of insurance.

19. Given the premium and rate of insurance to find face of policy.

> A man paid 50 dollars for insuring a house, the rate being 2 and ½ percent. What was the face of the policy?

Table XXIV gives the frequency of occurrence of each type in each of the eight texts examined. This table is to be read as follows: 10 problems classified as belonging to Type 1 which were found in Text 1; 30 such problems were found in Text 2; 20 in Text 3; 34 in Text 4, etc. The total number of problems classified under Type 1 is 147.

The variations in the frequency of the occurrence of problems belonging to a single type are worthy of notice. Some types have a high frequency in certain texts while in other texts their frequency is low and in many cases they do not occur at all. This means that different authors have tended to use different vocabularies.

TABLE XXIV. FREQUENCY OF OCCURRENCE OF TYPES OF PROBLEMS IN DIVISION IN EIGHT TEXTS

Type Number	Text								
	1	2	3	4	5	6	7	8	Total
1	10	30	20	34	8	13	13	19	147
2	28	14	123	11	-	75	49	27	327
3	-	7	6	2	4	11	3	15	48
4	11	24	24	2	19	20	14	17	131
5	2	4	2	-	3	1	-	1	13
6	5	10	4	9	3	13	4	52	100
7	-	2	2	3	2	1	1	2	13
8	1	1	-	4	-	4	1	-	11
9	2	7	3	-	1	1	2	12	28
10	-	1	13	2	2	13	3	2	36
11	6	19	13	3	-	4	11	16	72
12	1	-	2	-	1	1	3	-	8
13	-	1	1	2	4	1	-	1	10
14	9	42	12	41	3	14	39	30	190
15	-	12	5	-	1	-	18	4	40
16	1	2	1	1	1	8	1	8	23
17	6	1	1	1	-	2	2	-	13
18	8	-	4	-	1	-	3	7	23
19	9	-	3	1	-	-	1	1	15

Space does not permit the reproduction of similar tables for addition, subtraction, multiplication and the classes of two-step problems. In Table XXV a summary of the frequencies of the occurrence of the several types is given. This table is read as follows: In the case of problems requiring only addition, three types occurred in all eight texts, one occurred in seven and two in six texts. The total number of types occurring in five or more of the texts is six. The total number of problems classified in these six types is 464. The total number of problems is 622.

The reader should bear in mind that no attempt was made in this classification to determine what problems pupils should be asked to solve. The problems have been taken as they occurred in the texts. In effecting the classification, no consideration was given to the question of whether the problem was practical. In fact the purpose was not to obtain a list of practical problems but to secure a list of the forms of statement or language which had been used in the one- and two-step problems by the authors of widely used texts. Many of the technical terms of arithmetic are used (probably must be used) whether the problems are practical or not.

Experimental selection of problems. In order to secure data for the construction of a series of reasoning tests in arithmetic, about 300 problems were selected out of the total number examined and classified. Out of this number 156 problems were chosen for an experimental series of tests. In making the selections for this purpose the writer considered, in addition to the classification described above, the social importance of the problems. Thus a few types of problems which occur in a majority of the texts and have a high total frequency, were not represented. This introduces an additional subjective factor but in view of the emphasis which is being placed upon the social importance of the subject matter, the writer believes it is better to exercise judgment in this instance rather than to follow blindly statistics based upon the content of our present texts, particularly when it is obviously impossible to include representative problems of all types within a single series of tests of suitable length for classroom use.

TABLE XXV. FREQUENCY OF TYPES OF PROBLEMS

Operation	Number of Types occurring in				Total No. of Types	Frequency of problems classified	Frequency of all Problems
	8 texts	7 texts	6 texts	5 texts			
+	3	1	2	0	6	464	622
−	1	2	4	0	7	211	456
×	5	5	2	3	16	1641	1938
÷	5	6	3	5	19	1248	1610
+−	1	2	2		5	199	346
+×	4	2	3	4	13	472	718
+÷	1	0	1	3	5	127	299
−×	0	0	1	7	8	166	559
−÷	1	0	3		4	114	413
××	3	1	2	4	10	429	581
×÷	1	4	2	6	13	704	1234
++							7
÷÷							70
Total	25	23	25	32	105	5775	8853

In constructing the experimental series, Test I was designed for grades four and five, Test II for grades six and seven, and Test III for grade eight. Some such division is necessary because certain social situations from which problems are taken are not studied until the later grades although the mathematical relationships are very simple. Pupils cannot be expected to solve such problems until they are acquainted with the social situations. For this reason all problems involving percentage were placed in Test III. No consideration was given to the relative difficulty of the problems in making this division except that no problems requiring common fractions were placed in Test I and for the most part decimal fractions were confined to Test III.

Each test consisted of sixteen problems printed on a four page folder with space so that the pupil could do all of his work upon the test paper. The test papers showed that unless the pupil made errors and did his work over or used an elaborate method, ample space was provided except in a very few cases. The directions for administering the preliminary tests were essentially the same as those which now accompany the tests.

A number of cities were invited to cooperate by giving the tests between April 1 and 15, 1918. Fourteen cities responded, nine in Kansas, and one city in each of the following states: Illinois, Ohio, Michigan, New York, and Pennsylvania. Usable returns were received from 12,859 pupils.

When the record sheets and the test papers were returned to the writer it was found that the directions for marking the papers were not sufficiently complete and explicit. Consequently, there was a lack of uniformity in the marking. In order to insure uniformity the writer, assisted by two clerks rescored the papers. Whenever an unusual or questionable solution was found a record was made and all similar solutions were marked in the same way. In this way a high degree of uniformity in the marking of the papers was secured. Space does not permit a detailed statement of the plan of scoring of the solution of each problem but the general plan may be indicated.

The solution of a problem was considered correct in *principle* if the pupil's work showed that he had based his solution upon the relationships which exist between the quantities of the problem. For example, in the problem, "If a man has $275 in the bank and draws out $70, how much has he left in the bank?", there are *three* quantities: $275, $70, and the amount "left in the bank." These are related so that the difference between $275 and $70 must equal the amount left in the bank. A solution of the problem based upon this relationship must involve the subtraction of $70 from $275, or the finding of a number which added to $70 will make $275.

In the problem, "A house rents for $35 a month. This is how much a year?", the three quantities are $35, 12, or the number of months in a year, and the amount for a year. The relation is that the product of $35 and 12 equals the amount of rent for a year. A solution based upon this relationship would usually be one in which $35 was multiplied by 12. In a few cases the pupil had set down $35 twelve times and added. This solution was counted as correct in principle because it was considered that the pupil had recognized the relationship which existed between the quantities of the problem. Incidentally it should be noted that although such a solution was counted as being correct in scoring the papers of the test, a teacher should not encourage it. In fact the writer believes it should be discouraged, except possibly when the pupil is learning the idea of multiplication, because the method is not efficient. It requires more time and there are more opportunities for error in the mechanical work.

In the case of the above problem, if a multiplier other than 12 was used the solution was counted as correct in principle because it was considered that correct recall of denominate number facts was not a part of the reasoning. In a few cases 35 was multiplied by itself. This was marked incorrect in principle.

Although the pupils were directed to do all work upon the test papers a few gave only the answer in the case of certain problems. They had either solved the problem mentally or on another sheet of paper. An arbitrary rule was adopted. If the answer was correct the problem was marked correct in principle and answer. If the answer was wrong it was marked incorrect in both principle and answer.

An answer was not marked correct unless the solution of the problem was correct in principle and the answer was numerically correct and in its lowest terms if it contained a fraction. It was not required that the answer be labeled with its denomination.

Weighting the problems. For each problem three records were secured: (1) Number of pupils attempting the problem. (2) Number of solutions correct in principle. (3) Number of correct answers. From these facts the percent of solutions correct in principle and the percent of those solved according to the right principle which had also correct answers were calculated. These percents were translated into *sigma* values. The former being designated as the "P" value of the problem and the latter as the "C" value. In doing this it was assumed that the ability to solve problems was distributed normally and included between +2.5 *sigma* and —2.5 *sigma*. The tables given in Rugg's "Statistical Methods as Applied to Education" were used. The values were calculated to two decimal places but in order to simplify the computation of scores they were expressed in terms of the nearest integer in the tests as now published.

In the case of those problems which were solved by the pupils in two successive grades, the average inter-grade interval was found for each group of problems by taking the average of the differences of the *sigma* values of the problems of the test. This inter-grade interval was added to the values of the problems for the upper of the two grades to reduce them to the basis of the lower grade. The average of the two values was taken as the final value of the problem.

An attempt was made to reduce the *sigma* values to a common zero point, and thus secure comparable scores, by having a limited number of problems from Test I appear in Test II and also a limited number of problems from Test II appear in Test III. It happened that some of the problems chosen showed inversion and for this reason it was deemed advisable not to attempt to reduce the values to a common zero. Thus the scores obtained from the different tests of the series are not comparable.[7]

Construction of the final tests. Out of the 156 problems included in the preliminary test, 90 which belonged to types occurring in five or more of the eight texts examined, were selected for the final tests. Since in the selection on the basis of content there was no effort to include problems which exhibited wide range of difficulty, there was no attempt to construct a difficulty scale. In fact it is the judgment of the writer that the educational objectives implied by such a scale in the field of problem-solving are open to serious criticism. In our schools we should endeavor to instruct pupils to solve problems because they are socially worth while rather than because they exhibit a certain degree of difficulty. The purpose here is to construct a group of tests containing problems that are representative of the language in which problems are stated in our representative text books and which appear to be satisfactory for testing purposes.

The final tests consist of 15 problems each. Test I is for grades four and five, Test II for grades six and seven, and Test III for grade eight. There are two forms of each test. In selecting the 90 problems for these tests those were rejected which were commented on unfavorably by those who gave the preliminary tests. Also those problems were rejected which were found to be particularly confusing to pupils. The arrangement of the order of the problems in a test was made without reference to their difficulty values. An attempt was made to secure as high degree of variation in the operations required as possible. In the two forms of each test the corresponding problems are approximately equal in difficulty, and so far as possible, the two forms were made equivalent in other respects.[8]

---

[7] The method of weighting is open to criticism. It is used in an attempt to give more credit for doing a difficult problem than for doing an easy one. It is not at all certain that such a plan gives the most truthful indication of a pupil's ability. Some recent studies have shown that unweighted scores correlate very highly with the weighted scores obtained by this method. Therefore, it is likely that the tests would have been nearly as accurate measuring instruments without any determination of weights.

[8] No determination of the reliability or validity of these tests was made as a part of the original derivation. Neither is it possible to make a report on these questions at this time. Some work which was done on the question of reliability indicated that the tests were less reliable than tests in the operations of arithmetic and in silent reading. This appeared to be due to the fact that frequently pupils are unable to do certain problems because of a peculiar course of study.

Analysis of errors made by pupils on preliminary test. In the preliminary testing the following six problems were given to 100 fifth grade pupils in one city. The results of an analysis of the test papers are given in Table XXVI.

1. Mrs. Black received $2 a yd. for broadcloth. She sold 78 yds. How much did she receive?

2. At the store a towel roller costs 35c. George made one for his mother. He used 12c worth of lumber, 2c of hardware, and 3c worth of shellac. Find how much George saved his mother.

3. A Kansas farmer bought 80 acres of cheap land for $240. Oil being found on his farm he sold the land for $60,000. What was his profit?

4. A car contains 72,060 lbs. of wheat. How much is it worth at 87c a bushel?

5. A field is 20 rds. long and 12 rds. wide. How many rods of fence are needed to enclose it?

6. What are the average daily earnings of a boy who received 88 cents, 25 cents, $1.15, 75 cents, 50 cents, and 60 cents in one week?

TABLE XXVI. RESULTS OF ANALYZING THE ERRORS OF 100 FIFTH GRADE PUPILS

	PROBLEM						
	1	2	3	4	5	6	Total
Number of pupils attempting	100	100	92	52	94	94	532
Errors in reasoning	8	21	39	38	33	67	206
Errors in fundamentals	1	7	26	16	2	35	87
Omissions and errors in copying	2	0	2	3	0	4	11
Errors in decimals	13	0	16	25	0	14	68

Two significant facts are shown in this table. First, a majority of the errors (55 percent) are in reasoning. More than one-third of the attempts (39 percent) resulted in faulty reasoning. Second, 41 percent of the errors in calculation were in placing the decimal point. This second fact becomes more significant when we note that these errors occurred in problems involving only United States money and that the first and third problems which produced 29 of the 68 errors do not really involve decimal fractions. In these two problems the error consisted in pointing off the answer when it should not have been done.

The wide spread use of the Courtis Standard Research Tests, Series B and other tests upon fundamentals has resulted in increased attention to the fundamental operations with integers. There should be no decrease in the emphasis upon this phase of arithmetic for one out of six pupils made errors in the simple calculations required in these problems but the greatest source of error and therefore the greatest need is increased attention to the mental processes involved in reasoning as it occurs in solving arithmetical problems. The necessity for doing this becomes more obvious when we examine the nature of the errors in reasoning. In problem 2, 13 of the 21 errors in reasoning were due to adding all of the terms; in problem 4, 33 out of 38 of the errors in reasoning were due to multiplying 72,060 by 87 without attempting to reduce the pounds to bushels; in problem 6, 58 of the 67 pupils who reasoned incorrectly simply added the terms. Each of these errors may be ascribed to inaccurate or incomplete reading of the problem, or the first step in the rational solution of a problem.

In tabulating the scores of the preliminary test the variation in achievement of different classes was particularly noticeable. It was evident that some teachers were teaching their pupils to solve problems while others, frequently within the same school system, were not doing so. It is also significant that a number of teachers consistently marked as correct, solutions which were clearly incorrect. This may have been accidental on the part of the teacher but this is very doubtful. A striking illustration was furnished by the problem: "A baker used 3/5 lbs. of flour to a loaf of bread. How many loaves could he make from a barrel (196 lbs.) of flour?" The correct solution requires that 196 be *divided* by 3/5 which gives an answer of 326 2/3 loaves. In several instances teachers marked as correct all papers in a class in which 196 was multiplied by 3/5. This latter solution gives an answer of 117 3/5 loaves. The fact that teachers made such errors as this indicates that they are not familiar with solving reasoning problems. Perhaps this is one source of the poor records made by the pupils.

CHAPTER IV

MONROE'S TIMED SENTENCE SPELLING TESTS AND PUPIL'S ERRORS

How ability in spelling should be measured. In measuring ability in spelling by having a pupil spell words which are dictated in lists it is clear that the conditions under which the pupil spells the words are not the conditions under which he spells the words which he uses in writing themes, letters, and other school exercises. As a result we probably fail to obtain a measure of his "true spelling ability."

If the test words are embedded in sentences and the sentences written from dictation we approach more nearly normal spelling conditions because the pupil is writing connected words which have meaning. A still closer approximation appears to be secured by dictating the sentences at approximately the rate at which the pupil is accustomed to write. By thus causing the pupil to write at approximately his normal rate of writing, he does not have time to study over the spelling of words, and as a result we secure a record of spelling which is largely automatic. Under such conditions a pupil's attention is centered primarily upon writing and not upon his spelling. Of course, these conditions are not those under which the pupil normally spells words. The writing from dictation may be an unusual exercise for the pupil. Some pupils will be accustomed to write more slowly than the rate of dictation. This may tend to confuse them. To what extent these and possibly other factors prevent our obtaining a record of the "true spelling ability" of the pupil by a timed-dictation test we do not know. It appears, however, that a *Timed Sentence Spelling Test* is likely to yield a more valid measure of spelling ability than a list of words dictated separately.

The construction of Monroe's Timed Sentence Spelling Tests. In order to make easily available a timed sentence spelling test, the writer constructed a series of such tests, using test words chosen from appropriate columns of Ayres' Scale for Measuring the Ability in Spelling and basing the rate of dictation upon the measurements of the rate of handwriting of over six thousand Kansas school children. In order that the scores might have a high degree of reliability as measures of the spelling ability of individual pupils, fifty test words were used in each test. According to one study[9] the probable error of an individual score for a test of fifty words is less than 1.00 when the score is expressed as the percent of words spelled correctly. For a class of twenty-five or more pupils the probable error of the class score would be 0.2.

---

[9] Otis, A. S. "The reliability of spelling scales involving a 'deviation formula' for correlation." *School and Society*, 4, (November 11, 1916), 716-22. This study deals with the reliability of tests consisting of isolated words and it is possible that the results might not apply to a timed sentence spelling test.

For grades III and IV the test words were taken from Column M of Ayres' Scale for Measuring Ability in Spelling. For grades V and VI they were taken from Column Q and for Grades VII and VIII and the high school they were taken from Coulmans S, T, and U. The test for the fifth grade is reproduced. The sentences are to be dictated when the second hand of the watch reaches the position indicated in the left-hand margin.

In this test no test words come at the end of the sentences. Thus, the pupil who writes slowly will be much less likely to make low scores because he does not have time to complete writing the sentences. It should also be noted that all other words found in these sentences are easier to spell as shown by the Ayres Scale. It was thought advisable to allow time for dictating the sentences in addition to the actual writing. For this reason, the rate of writing for each school grade was increased by ten percent, sixty-six seconds instead of sixty being allowed for the number of letters which pupils commonly write in sixty seconds.

Tentative grade Norms. This series of timed sentence spelling tests was given in sixteen Kansas cities in April and May, 1917. During the school year of 1919-20 scores were reported from a number of cities. The grade medians for these two groups of cities and the norms given by Ayres are given in Table XXVII. In comparing the successive grades it must be remembered that the same test words were not used for all grades. One list of test words was used for grades III and IV, another for grades V and VI and still another for grades VII and VIII and for the high school.

The fact that the median scores in Table XXVII are materially below Ayres' norms indicates that a different type of spelling ability has been measured by the timed spelling sentence test than that measured by Ayres in constucting his scale. (Ayres had the words dictated in lists). This fact becomes more apparent when it is recalled that many of the cities which gave these tests had used Ayres' Scale as a minimum course of study as well as a source of test words. Thus, had the test words been dictated in lists it is likely that the m ian scores would have been materially above Ayres' norms.[10]

MONROE'S TIMED SENTENCE SPELLING TEST ARRANGED FOR THE FIFTH GRADE.

Seconds

60 The *president* gave *important information* to the men.
48 The *women* were *present* at the time.
19 The *entire region* was burned over.
49 The *gentlemen declare* the *result* was printed.
30 *Suppose* a *special attempt* is made.

---

[10] It is possible that the difference between the median scores and Ayres' norms may be due to factors other than the measurement of different types of spelling ability. Many of the pupils probably were not accustomed to writing from dictation and all were not accustomed to writing at the rate at which these tests were dictated. It is possible that these unusual conditions may have been operated to materially lower the scores of a number of pupils or even of most pupils.

60 The *final debates* were held.
24 The *factory employs* forty men.
51 *Sometimes* the *connection* is not made.
24 I *enclose* a *written statement* with the book.
3 *Prompt action* is needed.
25 It was a *wonderful surprise* to all.
55 The *addition* to the *property* was begun.
31 *Remember*, *Saturday* is the day.
57 They *await their* leader.
19 *Either* make another *effort* or return.
52 The *famous estate* is close.
16 In this *section* little *progress* was made.
53 The *measure* is *due* to pass.
16 A *position* in the *field* is his.
42 To *whom* was the *command* given?
8 *Whose claim* was bought?
29 He *represents* the *firm* in this matter.
2 Go *forward* in that *direction* to reach the city.

When the second hand reaches 43, stop the writing.

Allow no corrections or additions to be made. Ask the pupils to turn their papers over and write their name and grade. Appoint two or three pupils to collect the papers.

TABLE XXVII. MEDIAN SCORES (WORDS SPELLED CORRECTLY) AND AYRES' NORMS. TOTAL WORDS IN EACH TEST IS FIFTY

	Grade									
	III	IV	V	VI	VII	VIII	IX	X	XI	XII
1917 Testing										
Number of pupils	1342	1539	1427	1338	1136	876	289	251	188	96
Median (May)	28.6	39.4	32.8	41.8	35.1	43.0	43.3	45.1	46.5	47.9
1919-20 Testing										
Number of pupils	437	423	440	302	218	153				
Median	28.2	43.0	39.7	43.0	32.6	38.8				
Ayres Norms	33.0	42.0	36.5	42.0	39.0	44.0				

In Table XXVIII the distributions of the scores are given for grades III to VIII inclusive. In the fourth, sixth, and eighth grades there are a number of perfect scores which show that the test was not sufficiently difficult for the best spellers in these grades and, therefore, did not give us a measure of their spelling ability. This statement probably applies also to those pupils who spelled incorrectly one, two, or three words. From the standpoint of accurate measurement of the spelling ability of "good spellers," these tests are too easy, especially for grades four, six, and eight and the high school.

This table is also significant in another respect. The range in the number of words spelled correctly extends from zero, or at the most two, up to fifty and, as has been pointed out in the above paragraph, the spelling ability of some pupils probably extends beyond perfect scores on these tests. That the

TABLE XXVIII. DISTRIBUTION OF SPELLING SCORES ON MONROE'S TIMED SENTENCE SPELLING TESTS

Score	Grade					
	III	IV	V	VI	VII	VIII
50	7	45	13	40	14	33
49	11	53	23	83	21	48
48	13	82	29	75	31	64
47	17	76	44	81	45	60
46	24	83	30	82	40	55
45	28	89	38	77	38	61
44	17	63	50	78	39	59
43	34	64	38	81	54	59
42	36	58	49	60	34	48
41	37	45	49	62	30	40
40	35	76	55	69	37	33
39	39	57	31	47	40	25
38	29	67	35	48	34	29
37	30	42	55	43	30	28
36	43	53	44	34	44	17
35	45	43	42	38	40	25
34	43	52	41	23	29	22
33	27	36	40	32	40	18
32	30	43	46	20	31	19
31	42	30	40	24	41	12
30	35	33	46	33	29	24
29	35	33	41	16	23	13
28	31	34	35	17	23	12
27	32	44	29	16	23	10
26	30	23	40	13	28	7
25	34	16	39	18	32	14
24	37	23	42	9	23	3
23	38	26	27	15	13	10
22	26	19	16	3	27	5
21	30	21	30	13	14	5
20	39	10	25	11	18	4
19	23	9	21	5	24	4
18	22	6	24	7	14	1
17	21	6	16	9	18	2
16	30	9	18	9	21	1
15	24	13	22	6	20	
14	26	8	17	7	10	1
13	23	8	14	5	7	2
12	21	5	18	5	10	
11	22	2	15	4	8	1
10	23	8	9	2	8	
9	23	6	16	4	5	
8	25	1	9	2	4	
7	18	2	10	3	5	1
6	12	3	7	1	2	
5	10		6	3	6	
4	16		12		2	
3	11	4	11	2	4	
2	11	4	6	2	1	1
1	12	3	10	1	1	
0	15	3	4		1	
Total	1342	1539	1427	1338	1136	876
Median	28.6	39.4	32.8	41.8	35.1	43.0

pupils of a grade exhibit a wide range of ability is a well known fact. We also know that there is a very large degree of over-lapping between successive grades. Probably all would admit that a grouping of pupils for purposes of instruction exhibiting such great differences in spelling ability does not make for the most effective instruction. Correction of this condition is sometimes difficult. In ungraded schools or in rooms where there are two or more grades, it is relatively easy to group the children for spelling instruction, but in rooms where all the pupils belong to the same grade this is more difficult. The pupils who make perfect scores or scores nearly perfect might be excused from spelling but this would not take care of those most needing instruction.

In view of the fact that the time devoted to spelling is more frequently used in *testing* spelling than in *teaching* spelling, this table suggests that there is a definite need for teaching some pupils to spell. Some pupils appear to have learned to spell under the type of "instruction" which the schools now provide, but a large number of pupils have not done so. These pupils require a different type of "instruction." They probably need some assistance in learning to spell.

Spelling errors. In the 1917 testing, several cities returned, in addition to the class record sheets on which were recorded the scores, the test papers. The specific errors in spelling the test words have been carefully tabulated for 430 third grade pupils, 463 fourth grade pupils, 294 fifth grade pupils, 188 ninth grade pupils, 120 tenth grade pupils, 107 eleventh grade pupils, and 169 twelfth grade pupils. This study of errors was not extended as originally planned to the pupils in the sixth, seventh, and eighth grades because of a lack of funds.

The fundamental reason for giving tests is to secure information concerning the abilities of pupils. In order that this information may be most helpful to us it is necessary that we interpret the scores in terms of pupil needs. The fact that a pupil has a certain score and that this score is above standard or below standard becomes significant to the teacher only when this fact is expressed in terms of the learning needs of this pupil. In the case of scores below standards we have only partially interpreted them when we say that the pupils having such scores need to place more emphasis upon their spelling or that they need assistance from the teacher. We carry the interpretation a step further when we specify what words the pupils need to give their attention to. We may carry our interpretation another step by ascertaining the particular errors which the pupils are making and hence need to correct.

The tabulations of the misspellings show that some ways of misspelling a word occur much more frequently than others. In general about 60 percent of the misspellings belong to one of five kinds and about 30 percent are included in one. If a teacher knows the types of misspellings which are most likely to

TABLE XXIX. THE FREQUENCY OF MISSPELLINGS OF THE WORDS "COLLECT" AND "OMIT" AS FOUND IN THE TEST PAPERS 463 FOURTH GRADE PUPILS. WHERE NO FREQUENCY IS GIVEN THE MISSPELLING OCCURRED ONLY ONCE.

Collect			*Omit*		
colect	110	culate	omite	35	poment
clect	21	cucket	oment	22	phone
cleck	9	colest	omet	19	omoit
clet	5	connect	omitt	10	amint
celect	4	cluck	omeat	10	only me
clack	4	polet	ommit	8	eneat
clocket	3	coiker	omitte	6	anete
colcet	3	cluct	onit	5	enit
cloct	3	coulack	ownit	5	bremet
cleat	3	cleick	admit	5	obeat
cluk	3	clach	omnit	3	ument
colet	2	clacket	ament	3	opit
col	2	klechit	omnitt	3	omett
collet	2	clich	omt	2	oumient
coloct	2	clank	amit	2	omemt
corlect	2	klocet	ameit	2	portment
cllect	2	klucklect	amite	2	owmet
clock	2	clouct	omate	2	oneit
culect	2	colict	omeet	2	omant
cocet	2	colcailate	only	2	owe meat
comelet		clloe	onipe		connect
ceaclect		caluct	omot		obment
calect		clearty	amite		pomit
somelet		coloce	omiet		onitt
golet		colack	poemet		nent
klited		claxt	omitted		inate
dontlet		conckle	onight		oak
kleke		conut	anitt		onight
kolicket		coulet	abment		own
cleakely		collice	onet		
cault		clecet	comen		
celec		colcect	onent		
clecty		cacated	emmit		
codect		cleclect	amited		
colext		clucker	otamate		
collete		clecket	offit		
clllect		cleack	adnet		
collecks		colucket	aneat		
clete		concect	anint		
cole		cart	omoitte		
cleatlet		lecke	opote		
cherd		coul	ohmit		
colectt		coleck	omote		
cocllict			obed		
kalat			omip		
clectket			amet		
cleakit			tell me		
coluct			anpend		

occur she can, in her teaching of spelling, warn pupils against the errors which they will most likely make and give them such training as may be required to insure that they will not make these errors.

Because of the importance of giving attention to spelling errors, we give in Table XXIX the various misspellings of the words "collect" and "omit" found in the examination of 463 fourth grade test papers. Some of the misspellings which occur only once or at the most two or three times probably should not be counted as "true" misspellings. The writing of c-o-l probably should be counted as an incomplete spelling. The writing of c-h-e-r-d indicates either that the pupil is not acquainted with the word "collect" or was unable to recall automatically the spelling, or failed to understand the word in the teacher's dictation. On the other hand the misspellings which recur frequently may be considered as "true" misspellings and are the ones which should be guarded against in the teaching of spelling.

In interpreting spelling scores, we need to consider the questions, "Are all misspellings to be treated alike? Does one misspelling mean the same

TABLE XXX. SHOWING MOST FREQUENT MISSPELLINGS OF THE TEST WORDS IN GRADE III BY 430 PUPILS

Correct form of word	No. of misspellings	Five most frequent misspellings									
		1		2		3		4		5	
account	218	acount	63	count	46	ocount	10	mony	8	accont	5
again	177	agin	52	agan	40	agen	30	agian	14	agun	3
almost	157	allmost	101	alnost	13	allnost	12	ailmast	3	amost	2
anyway	105	enyway	27	eneyway	7	inway	5	annyway	4	inyway	4
army	115	armey	19	arny	15	arme	15	armer	5	armay	4
begin	169	began	48	begain	36	begun	22	begen	16	begine	6
begun	196	began	49	begin	37	begain	30	begone	9	begon	9
beside	89	by	29	besid	20	becide	5	besaid	3	besied	2
both	72	those	21	bout	6	bothe	3	boath	3	bot	3
bought	185	bot	85	bout	19	brought	13	bough	11	boght	6
change	165	chang	23	changs	14	chain	11	chanes	10	chanse	6
children	135	childern	25	childen	12	chirdlren	5	childred	5	childer	3
collect	219	colect	36	clect	26	clact	4	clet	4	clack	3
contract	172	contrack	49	contrach	5	contrect	4	contracked	4	contrak	4
deal	167	deel	45	dill	28	dell	27	dile	17	del	6
died	120	dide	50	did	18	dided	13	dead	7	diead	2
dress	55	drees	12	dreess	8	drass	8	dres	2	dreass	2
drill	69	dril	21	drell	7	dill	5	grill	4	drail	3
driven	77	drive	11	driving	5	dreven	5	drivin	5	drivan	4
enter	159	inter	71	anter	13	enture	7	intre	4	ender	4
extra	174	extry	39	exter	10	antra	5	extery	4	extray	4
few	134	you	48	fue	32	to	8	flew	6	do	4
follow	123	folow	14	foller	8	fallow	7	follo	6	flow	6
goes	162	gose	48	gos	47	go	44	goese	2	goe	2
great	188	grate	91	grat	33	graet	10	grait	8	grad	3
income	140	incom	12	incone	8	come	5	encome	2	incume	1
inform	107	inforn	17	reform	12	infrom	9	imform	6	informe	6
members	122	menbers	18	meners	9	mambers	8	mebers	5	mimbers	4
might	155	mite	78	mit	24	mint	11	night	5	mighe	3
money	126	mony	63	many	9	maney	8	noney	3	momy	3
month	149	mounth	42	moth	30	mouth	17	mount	9	mont	7
office	138	ofice	23	offic	21	offes	16	offer	8	ofis	7
omit	188	omet	25	omite	21	oment	15	onit	14	omeat	11
paid	167	pade	51	payed	32	pad	28	pay	18	payd	5
past	119	passed	30	pass	22	pas	9	pasted	7	pask	4
picture	131	pitcher	17	picher	13	pictur	9	pictors	7	pctcure	7
please	165	pleas	77	plese	17	ples	8	pleace	7	place	6
provide	128	provied	10	provid	8	proved	8	proveid	7	prove	5
railroad	111	railrode	12	rayroad	7	ralroad	4	railroat	3	roilroad	3
ready	144	redy	36	rady	22	read	9	redey	7	rade	5
recover	120	cover	22	uncover	20	recove	9	recuver	8	recome	3
return	98	reture	13	returne	7	retun	5	returne	4	retern	4
says	138	sais	19	said	17	say	13	saids	10	sas	9
those	140	thoes	76	thos	14	thouse	6	thous	5	thoses	4
ticket	69	tick	18	ticked	5	tickit	3	ticke	3	tecket	3
took	88	tuck	16	tock	13	take	7	tooke	5	tuk	5
unable	82	unabel	10	onabel	6	unabil	5	unble	5	onable	5
understand	83	unstand	13	under	5	understant	5	undestand	4	understend	3
while	109	whil	22	wile	21	whill	11	will	10	why	6
who	124	how	48	ho	35	hoe	8	hew	6	hoo	2
Total	6743		1845		903		544		331		221
Percent			27		13		8		5		3

as any other misspelling?” Does “colect” have the same meaning with reference to the pupil's need for instruction as “connect” or “cole” or “ceaclect?” Probably all misspellings do not indicate the same instructional needs. The writing of “colect,” as was done by nearly one-fourth of all these fourth grade pupils, indicates uncertainty concerning the double consonant or a wrong habit formed. “Connect” probably indicates that the word was misunderstood. “Ceaclect” indicates the lack of a fixed habit and possibly that the pupil was not acquainted with the word.

In Tables XXX, XXXI, and XXXII, we give a list of the five most frequent misspellings[11] of the test words in grades III, IV and V. The column

TABLE XXXI. THE MOST FREQUENT MISSPELLINGS OF THE TEST WORDS IN GRADE IV BY 463 PUPILS

Correct form of word	No. of misspellings	Fivemost frequent misspellings									
		1		2		3		4		5	
account	246	acount	104	count	42	ocount	12	accont	7	acont	7
again	101	agin	40	agan	27	agian	9	agen	9	agine	2
almost	136	allmost	83	alnost	28	most	8	alnose	2	almos	1
anyway	76	enyway	29	eneyway	7	any	6	inyway	5	anway	2
army	78	armey	17	arny	9	arm	7	arme	4	arney	4
begin	137	began	52	begain	33	begen	16	begun	9	begon	2
begun	200	began	59	begin	33	begain	29	begon	10	becine	8
beside	55	besid	21	by	9	becide	5	besde	2	aside	2
both	77	bouth	16	those	14	bough	10	bothe	6	bought	5
change	144	chang	21	changes	11	changed	8	chanes	8	chane	8
children	63	chirldren	11	childern	6	chirlden	5	chrildren	4	chidern	3
collect	257	colect	110	clect	21	cleck	9	clet	5	celect	4
contract	178	contrack	48	comtract	25	contrat	8	concrat	7	contrach	6
deal	106	deel	26	dell	24	del	6	dill	6	dile	4
died	72	dide	29	die	9	dided	6	did	6	Died	3
dress	31	drees	7	dreess	3	dres	3	Dress	2	drese	2
drill	59	dril	10	drell	7	dill	4	trill	3	Drill	2
driven	66	drivin	9	dreven	4	drive	8	drivvin	3	driver	3
enter	99	inter	61	enten	3	entere	3	anter	3	ender	3
extra	159	extry	45	extray	8	extery	6	exter	5	axtra	4
few	132	you	56	fue	32	fiew	5	tue	3	fhew	2
follow	97	fallow	19	flow	11	folow	10	fllow	6	foller	5
goes	88	go	34	gose	30	gos	17	gois	2	was	2
great	95	grate	62	geart	5	grat	4	greate	4	gred	3
income	90	incone	18	incom	16	encome	7	incon	4	icome	3
inform	132	inforn	33	imform	16	informe	13	enform	7	enforn	6
members	129	menbers	31	membors	7	memers	6	meners	5	membes	3
might	121	mite	57	mit	14	mint	14	migh	5	night	4
money	79	mony	48	noney	7	nony	2	momey	2	moeny	2
month	114	mounth	31	mouth	18	nounth	11	moth	9	mount	5
office	94	ofice	20	ofic	15	offic	6	offer	6	offece	5
omit	205	omite	35	oment	22	omet	19	omitt	10	omeat	10
paid	103	payed	50	pade	15	pad	12	paied	6	Paid	2
past	69	passed	25	pass	18	pas	5	pasted	4	pass	2
picture	75	pitcher	16	pitcure	10	Picture	3	pictur	3	pictor	3
please	121	pleas	66	plese	14	plase	5	plise	5	pleace	4
provide	160	provied	19	provid	12	pervide	8	proved	5	porvide	4
railroad	73	railrode	21	ralrode	3	railrad	3	roilroad	2	Rail Road	2
ready	78	redy	27	rady	19	reddy	6	read	3	raid	2
recover	118	cover	40	uncover	17	recove	12	recuver	6	rncover	3
return	62	returne	11	retun	8	retern	8	retrun	4	retune	3
says	105	said	24	sais	16	say	4	ses	6	saids	6
those	96	thoes	58	thos	10	thouse	4	thoese	3	thoses	3
ticket	46	tick	5	tichet	5	ticet	3	tictect	2	tickt	2
took	47	tuck	10	tock	8	toke	7	tuke	2	take	2
unable	73	anable	16	onable	8	unabel	5	inable	5	uable	5
understand	61	unstand	13	understad	5	undersand	4	understan	3	undestand	3
while	61	whil	16	will	8	whill	7	wile	5	whal	5
who	69	how	26	ho	22	Hew	5	wow	3	whow	2
Total	5133		1683		714		383		236		173
Per Cent			33		14		7		5		3

headed "number of misspellings" gives the total number of misspellings. In the following columns the five most frequent misspellings are given with their frequencies.

The column giving the total number of misspellings shows clearly that these words were not equal in difficulty for these pupils, although they were taken from the same columns of Ayres' Scale. This condition indicates a limitation of that scale. This may be due to the fact that in deriving the scale Ayres used scores from many different states and hence the evaluation of the words was more general. It may also be due to the use of the scale as a minimum essential list. A third reason may be the nature of the test which was used.

TABLE XXXII. THE MOST FREQUENT MISSPELLINGS OF TEST WORDS IN GRADE V, BY 294 PUPILS

Correct form of word	No. of misspellings	Five most frequent misspellings									
		1		2		3		4		5	
action	35	actin	3	acionn	3	axion	1	actshon	1	axchian	1
addition	127	addion	24	adition	21	addation	8	addtion	8	additon	8
attempt	144	atempt	34	attemp	12	attenpt	11	etempt	6	atemp	6
await	136	awake	62	awate	18	awaked	13	wait	8	awaite	4
claim	145	clame	79	plain	12	clam	12	clain	11	claime	6
command	111	comand	62	comman	5	commend	4	commad	4	comend	4
connection	95	conection	47	conecton	4	coniction	3	conections	2	connect	2
declare	152	declared	65	declair	17	declaired	7	declar	6	declard	5
due	78	dew	35	do	29	doe	3	du	2	doo	2
effort	122	effert	28	efert	21	efort	15	effer	6	effor	5
Either	155	Eather	99	Ether	20	Eathe	3	either	2	Eaither	2
employs	178	imploys	43	employes	17	inploys	13	imployes	10	employ	7
entire	123	intire	80	inter	7	intier	6	intiar	3	intre	2
estate	99	astate	41	state	23	asstate	4	estait	2	states	2
direction	84	dirrection	14	drection	10	way	7	derection	5	driction	3
famous	109	famious	27	famis	10	famos	7	fames	5	famest	5
field	33	feild	22	fields	2	feeld	1	feilds	1	fieled	1
firm	69	ferm	17	firn	11	furm	9	fern	8	form	3
forward	74	foward	28	foreward	14	frward	3	fored	1	forwar	1
factory	72	factor	13	facture	8	factry	4	fatory	3	factiory	3
final	88	finel	34	finial	11	finally	5	fineal	3	finall	3
gentlemen	113	gentleman	41	Gentlemen	9	gentle	5	gentlmen	4	gentlenen	3
important	117	importent	18	inportant	7	inportent	7	importanc	6	imporant	6
information	129	imformation	32	infermation	7	message	5	imfermation	5	infomation	3
measure	55	masure	24	mesure	3	maisure	2	meashure	1	maisure	1
present	57	presant	13	preasant	3	preasent	3	prisent	2	presint	2
president	131	President	33	presedent	9	Presdent	6	Presedent	4	presedient	3
Prompt	123	Promp	40	prompt	13	Promt	7	Promped	6	pronpt	4
property	89	propty	24	proptery	9	propity	6	propety	5	propery	4
progress	101	progus	9	progess	8	progres	8	prograss	5	progrees	3
position	289	possion	42	postion	34	posion	25	physician	9	possition	7
region	137	regin	20	regon	20	reigon	12	reagon	12	regen	7
Remember	156	Rember	93	rember	12	Remeber	8	Remmber	6	remember	4
represents	132	repesents	16	repersents	13	represent	13	repersent	4	reperzents	3
result	30	rezult	4	reselt	3	ruselt	2	resuled	2	it	2
Saturday	123	saturday	37	Saterday	27	Sat.	5	Saturdy	5	Satuarday	5
section	54	sextion	9	secton	6	sexion	2	sexton	2	Section	1
statement	97	statment	41	stament	5	statemet	4	satment	3	statemen	3
special	104	specil	11	specal	11	spechal	3	speshel	3	speacle	3
Suppose	188	Supose	67	suppose	56	Sopose	19	Soppose	9	supose	3
surprise	183	suprise	103	supprise	26	suprize	7	surprize	6	supprize	3
Sometimes	46	Sometine	13	Sometines	6	Sometine	5	Sometime	3	Sontimes	3
whom	96	shome	17	hom	14	home	13	who	9	hoom	9
Whose	127	Whos	31	Sho's	23	Who	21	Whoes	16	whos	7
women	83	wemon	24	wemen	13	woman	11	weman	10	wimen	4
wonderful	86	wounderful	23	wonder	8	wondful	8	woundful	15	wonderfull	5
TOTAL	5075		1642		619		346		239		173
PER CENT			32		12		7		5		3

---

[11] This includes in certain cases, the wrong use of the capital letter. A test word beginning a sentence was counted wrong if not capitalized. A test word which should be begun with a capital letter was counted wrong if the pupil failed to do this.

APPENDIX

In the following tables, 5, 10, 20 30, 40, 50, 60, 70, 80, 90 and 95 percentile scores are given for several tests. In general these are tests which have been widely used and which appear to have some permanence. The interpretation of a percentile score is similiar to that of a median. In fact the 50-percentile score is the median. The 20-percentile score is the score above which there are eighty percent of the scores and below which there are twenty per cent. These tables are to be used for determining the position which a pupil occupies in the total distribution of his grade.

TABLE I. MONROE'S STANDARDIZED REASONING TESTS IN ARITHMETIC FORM I. PERCENTILE SCORES BASED ON APRIL TESTING

CORRECT ANSWER

Percentile	GRADE				
	IV	V	VI	VII	VIII
95	16.5	21.1	20.0	23.5	18.4
90	14.4	19.1	17.8	21.2	16.5
80	11.7	16.5	14.9	18.5	14.0
70	9.9	14.7	13.2	16.5	12.1
60	8.4	13.0	11.7	14.9	10.4
50	7.0	11.3	10.4	13.4	9.0
40	5.9	9.7	9.0	11.9	7.5
30	4.7	8.0	7.6	10.4	5.9
20	3.5	6.2	6.0	8.5	4.2
10	2.0	4.0	4.0	6.0	2.4
5	1.1	2.4	2.8	4.0	1.4
No. of pupils	2968	2996	3518	2803	2515
CORRECT PRINCIPLE					
95	26.7	33.8	26.5	28.9	29.6
90	23.0	31.5	23.8	27.9	26.6
80	18.5	27.5	20.7	25.5	24.0
70	15.6	25.7	18.2	23.2	21.8
60	13.3	22.2	16.0	21.4	19.5
50	11.3	19.2	14.2	19.7	17.2
40	9.2	16.4	12.7	17.7	15.1
30	7.2	13.7	11.1	15.4	12.7
20	5.2	10.3	9.2	13.0	9.9
10	2.8	6.4	6.6	9.9	6.4
5	1.1	3.9	4.7	7.5	4.1
No. of pupils	2932	3027	3498	2706	2472
RATE					
95	16.4	23.4	18.1	20.4	17.3
90	14.5	19.5	15.5	18.3	14.9
80	11.8	15.9	12.9	15.4	12.0
70	10.0	14.0	11.4	13.7	10.1
60	8.7	12.3	10.0	12.3	8.6
50	7.8	11.2	8.7	11.2	7.5
40	6.7	9.9	7.9	9.9	6.6
30	5.7	8.6	7.0	8.6	5.7
20	4.5	7.4	5.8	7.6	4.7
10	3.0	5.4	4.2	6.1	3.2
5	1.7	3.8	3.0	5.1	1.8
No. of pupils	1412	1705	1699	1717	1642

TABLE II-A. MONROE'S STANDARDIZED SILENT READING TEST. COMPREHENSION. PERCENTILE SCORES BASED OM MAY TESTING.

Form 1

Percentile	Grade									
	III	IV	V	VI	VII	VIII	IX	X	VI	XII
95	20.8	24.2	33.8	38.6	39.4	43.9	51.5	50.1	51.8	58.4
90	14.4	20.0	26.7	34.0	35.7	40.4	43.3	45.0	42.1	51.7
80	11.2	16.5	22.6	28.3	30.1	35.4	36.3	39.0	40.0	42.5
70	8.9	14.0	19.4	24.2	26.5	31.6	31.0	33.9	35.9	40.2
60	7.0	12.2	17.1	20.2	23.6	27.5	27.3	29.8	32.1	36.5
50	5.2	10.5	14.9	18.8	21.0	24.1	24.0	25.8	28.5	33.3
40	4.1	8.7	13.2	16.7	18.8	23.6	20.8	21.4	26.0	30.5
30	3.1	7.0	11.4	14.5	16.8	19.4	17.6	19.0	23.2	25.1
20	2.0	5.1	9.4	12.0	14.6	16.7	14.3	16.1	20.2	21.7
10	1.0	2.6	6.2	9.0	11.4	13.1	10.7	10.3	15.3	16.1
5	.5	1.3	4.0	6.1	9.4	10.8	7.4	8.2	11.6	11.3
Number of pupils	1464	1805	1729	1871	1270	1337	783	397	236	230

Form 2

Percentile	Grade									
	III	IV	V	VI	VII	VIII	IX	X	XI	XII
95	21.0	27.2	29.7	39.6	43.8	45.8	48.6	52.4	56.3	58.9
90	17.6	24.1	27.2	36.0	40.0	42.9	41.7	46.1	49.7	51.8
80	13.6	20.3	24.6	31.1	35.4	38.5	34.9	39.1	42.9	43.6
70	11.0	17.5	22.3	28.1	32.2	35.2	30.8	34.0	38.0	39.7
60	9.2	14.9	20.0	25.6	29.3	32.6	27.5	30.1	33.1	35.9
50	7.9	13.4	17.9	22.7	27.0	29.9	24.5	26.7	29.4	32.1
40	6.4	12.0	15.7	19.7	24.6	27.6	21.5	23.5	25.9	28.4
30	5.0	10.5	13.5	16.9	21.4	25.3	18.2	20.6	22.8	24.9
20	3.3	8.5	11.6	14.0	17.9	21.6	14.6	16.7	19.9	20.3
10	1.7	6.0	8.0	12.2	14.1	17.1	10.2	12.3	15.1	16.1
5	.8	4.2	6.5	8.0	11.3	14.0	6.8	10.0	11.3	12.6
Number of pupils	8741	10625	10157	9404	8791	7561	781	625	443	320

Form 3

Percentile	Grade					
	III	IV	V	VI	VI	VII
95	22.2	25.1	27.7	41.9	44.6	46.0
90	19.4	23.8	25.6	37.7	41.7	43.6
80	15.7	21.3	23.6	31.9	36.9	39.9
70	13.4	19.1	22.2	27.6	32.9	37.0
60	11.5	17.1	20.6	24.8	29.4	35.3
50	9.6	15.2	19.0	22.1	26.5	31.1
40	8.0	13.5	17.3	19.4	23.8	28.3
30	6.4	11.2	15.3	16.6	20.4	25.5
20	4.7	9.0	12.8	13.8	18.2	22.1
10	2.3	5.8	10.1	11.0	13.7	17.5
5	1.2	3.3	7.1	8.5	11.3	13.9
Number of pupils	1604	1680	1704	1580	1513	1186

TABLE II-B. MONROE'S STANDARDIZED SILENT READING TEST. RATE PERCENTILE SCORES BASED ON MAY TESTING

FORM 1

PERCENTILE	GRADE									
	III	IV	V	VI	VII	VIII	IX	X	XI	XII
95	98.0	111.0	143.5	145.4	146.6	148.9	136.6	143.0	138.3	152.1
90	84.5	105.9	126.9	143.8	142.7	146.9	120.9	130.1	130.9	137.4
80	70.7	81.0	110.0	115.5	120.3	142.7	109.5	120.2	121.5	129.4
70	60.6	80.6	96.4	103.6	110.9	135.6	100.8	100.4	105.0	122.7
60	52.2	72.8	87.0	95.3	102.0	115.1	86.7	88.5	92.0	107.6
50	44.5	65.2	78.8	89.0	95.1	105.6	82.6	84.4	86.6	102.7
40	38.0	58.1	73.0	73.0	88.8	97.2	78.1	80.3	83.1	88.2
30	33.0	51.7	65.9	67.6	82.8	91.1	73.3	75.6	79.5	84.2
20	20.6	44.4	56.7	60.2	66.6	82.5	66.3	70.7	74.4	80.2
10	10.6	34.3	44.6	49.8	58.2	62.3	55.3	61.5	67.9	69.2
5	5.2	25.0	34.6	41.1	51.5	53.4	50.0	54.9	58.3	60.8
Number of pupils	1464	1840	1828	1869	1279	1357	785	397	235	220

FORM 2

PERCENTILE	GRADE									
	III	IV	V	VI	VII	VIII	IX	X	XI	XII
95	117.4	126.7	128.4	165.9	167.2	168.0	137.9	138.7	140.4	147.9
90	98.5	122.2	126.3	161.7	164.4	165.9	130.4	132.2	134.1	139.2
80	78.0	107.0	122.1	147.0	157.3	161.8	123.6	123.6	125.8	130.7
70	71.8	96.9	114.3	136.3	146.0	154.6	107.6	107.2	109.8	123.5
60	67.0	90.7	104.6	124.3	137.4	145.2	101.5	103.2	104.1	107.5
50	62.6	76.6	97.2	114.1	131.5	137.4	87.2	88.6	89.0	102.3
40	55.7	71.9	91.5	107.8	115.8	131.5	82.9	83.8	84.3	87.6
30	45.5	67.1	76.9	100.8	109.1	115.5	78.1	79.0	79.6	83.1
20	36.8	62.2	71.3	82.7	101.2	107.7	72.0	74.1	73.1	77.5
10	28.9	49.5	63.1	72.1	80.5	88.1	62.9	67.1	61.9	69.6
5	15.5	37.8	55.4	54.9	72.2	79.7	56.1	57.2	54.7	59.8
Number of pupils	9739	10579	10150	9853	8767	7681	781	626	443	322

FORM 3

PERCENTILE	GRADE					
	III	IV	V	VI	VII	VIII
95	123.3	127.0	128.3	146.7	148.3	148.8
90	109.9	123.9	126.7	142.8	146.2	147.2
80	99.4	114.7	123.3	124.1	141.9	143.9
70	89.0	106.6	119.9	114.9	124.6	140.5
60	84.5	101.6	109.6	107.7	115.9	124.0
50	80.0	93.8	105.6	101.1	109.0	117.3
40	72.3	86.9	101.6	94.1	102.5	112.2
30	55.1	81.4	94.2	86.9	95.0	106.0
20	46.7	72.6	85.7	79.6	86.8	98.3
10	42.7	48.9	75.6	69.1	77.5	85.2
5	22.7	43.8	59.6	59.2	65.3	76.5
Number of pupils	1608	1695	1702	1579	1528	1179

TABLE III-A. CHARTERS' DIAGNOSTIC LANGUAGE TESTS FOR GRADES III TO VIII. PRONOUNS. FORM I. PERCENTILE SCORES BASED ON MARCH TESTING.

Percentile	Grade					
	III	IV	V	VI	VII	VIII
95	27.9	27.7	30.7	32.2	35.4	38.9
90	24.8	24.8	27.5	29.7	33.3	37.5
80	21.3	21.5	23.6	26.9	30.6	35.0
70	18.5	19.3	21.7	24.7	28.4	32.9
60	15.8	16.9	20.1	22.9	26.3	31.1
50	13.6	15.1	18.5	21.4	24.5	29.0
40	11.7	13.5	16.8	19.7	22.5	26.9
30	9.9	11.9	15.2	17.9	20.7	24.4
20	7.8	10.0	13.3	16.0	18.5	21.3
10	5.1	7.6	10.9	13.2	15.5	16.0
5	3.3	5.8	9.5	10.6	13.1	10.9
No. of pupils	787	864	895	1344	1566	1253

TABLE III-B. CHARTERS' DIAGNOSTIC LANGUAGE TEST FOR GRADES III TO VIII. VERBS (FORMERLY VERBS A.) FORM I. PERCENTILE SCORES BASED ON MARCH TESTING.

Percentile	Grade					
	III	IV	V	VI	VII	VIII
95	29.2	32.0	34.4	37.0	36.3	39.4
90	25.3	29.0	32.2	34.3	34.7	38.4
80	20.5	24.3	29.5	31.0	32.7	36.8
70	17.8	21.6	27.4	28.1	31.1	35.5
60	15.0	19.5	25.4	26.1	29.4	34.2
50	12.6	17.7	22.6	24.3	27.7	32.8
40	10.8	15.9	20.7	22.4	25.6	31.2
30	8.2	13.9	18.6	20.2	23.4	29.6
20	6.2	11.7	15.3	17.7	21.5	27.2
10	3.9	8.8	11.5	13.6	18.6	23.4
5	2.7	6.4	7.4	11.3	16.4	20.8
No. of pupils	365	403	373	478	539	638

TABLE III-C. CHARTERS' DIAGNOSTIC LANGUAGE TEST FOR GRADES III TO VIII. MISCELLANEOUS A (FORMERLY MISCELLANEOUS) FORM I. PERCENTILE SCORES BASED ON MARCH TESTING

Percentile	Grade					
	III	IV	V	VI	VII	VIII
95	27.9	23.8	22.3	29.7	31.9	34.6
90	24.1	18.6	19.1	26.6	28.9	31.6
80	15.9	15.1	16.9	23.0	25.8	28.3
70	11.5	12.6	15.0	20.6	21.9	26.2
60	8.7	10.8	13.3	18.6	19.7	24.1
50	6.7	9.3	11.6	16.5	18.9	22.3
40	5.7	7.9	10.2	14.6	17.2	20.3
30	4.6	6.4	9.0	12.7	15.5	17.9
20	3.3	5.1	7.3	10.8	13.4	15.4
10	2.0	3.4	5.5	8.5	10.4	12.0
5	1.4	2.6	4.1	6.5	7.9	9.1
No. of pupils	386	669	668	845	758	494

TABLE III-D. CHARTERS' DIAGNOSTIC LANGUAGE TEST FOR GRADES III TO VIII. MISCELLANEOUS B (FORMERLY VERBS B). FORM I. PERCENTILE SCORES BASED ON MARCH TESTING

Percentile	Grade					
	III	IV	V	VI	VII	VIII
95	27.5	31.7	33.7	38.7	39.0	38.5
90	20.2	29.5	31.5	36.4	37.2	37.3
80	16.7	25.9	28.8	33.5	34.6	35.7
70	13.3	22.9	26.4	31.4	32.8	34.6
60	10.3	20.2	24.0	29.4	31.2	33.4
50	7.9	17.8	22.0	27.3	29.4	32.0
40	5.0	14.6	19.5	24.7	27.4	30.7
30	3.7	11.9	16.9	21.7	25.4	28.9
20	2.5	9.4	14.4	18.0	21.9	26.7
10	1.3	6.5	10.9	13.9	18.2	24.3
5	0.7	3.7	7.1	10.2	15.5	21.2
No. of pupils	230	430	307	475	412	294

TABLE IV. CHARTER'S DIAGNOSTIC LANGUAGE AND GRAMMAR TEST FOR GRADES VII AND VIII. FORM I. PERCENTILE SCORES BASED ON APRIL TESTING

Percentile	Pronouns		Verbs		Miscellaneous	
	Grade		Grade		Grade	
	VII	VIII	VII	VIII	VII	VIII
95	27.9	33.4	36.4	34.5	20.9	30.6
90	23.5	32.5	34.1	31.5	17.0	26.4
80	17.6	27.7	27.4	26.3	12.9	21.7
70	13.0	24.0	17.6	21.3	10.4	17.0
60	10.2	20.5	11.8	16.8	7.8	14.4
50	8.0	17.1	7.8	14.0	6.3	11.9
40	6.3	13.7	5.0	10.9	5.1	9.5
30	4.9	10.2	3.5	8.3	3.3	7.2
20	3.5	7.0	2.3	5.4	2.3	5.1
10	2.1	4.2	1.1	3.0	.9	3.1
5	1.3	3.0	.5	1.9	.5	2.2
No. of pupils	936	657	434	497	332	362

TABLE V-A. WILLING'S SCALE FOR MEASURING WRITTEN COMPOSITION, STORY VALUE. PERCENTILE SCORES BASED ON MARCH TESTING

Percentile	Grade					
	IV	V	VI	VII	VIII	IX
95	76.3	89.9	93.2	95.7	96.1	95.3
90	66.1	85.5	88.2	91.5	96.0	90.8
80	57.5	77.7	81.5	88.1	88.6	87.5
70	52.1	71.3	76.8	82.8	84.6	84.9
60	46.8	65.0	72.6	78.0	80.5	82.2
50	41.5	58.7	68.1	74.0	76.6	79.0
40	36.9	52.8	63.3	69.2	72.9	73.8
30	32.7	46.6	58.0	63.5	68.7	68.4
20	28.4	40.2	51.5	50.1	62.2	62.3
10	24.2	31.5	43.1	46.0	54.7	53.3
5	22.1	26.0	37.0	37.4	46.9	43.6
No. of pupils	320	579	705	692	572	129

TABLE V-B. WILLING'S SCALE FOR MEASURING WRITTEN COMPOSITION, ERRORS PER 100 WORDS. PERCENTILE SCORES BASED ON MARCH TESTING

Percentile	Grade					
	IV	V	VI	VII	VIII	IX
95	32.1	28.8	18.8	18.1	14.8	11.9
90	31.7	24.9	16.0	13.9	11.9	10.3
80	28.2	19.4	11.9	10.8	8.8	7.9
70	23.9	15.6	9.9	8.7	7.4	6.0
60	20.9	12.5	8.2	7.3	6.1	5.2
50	18.5	10.7	6.8	5.8	5.2	4.4
40	15.5	9.3	5.5	4.8	4.4	3.6
30	13.0	7.3	4.2	3.8	3.6	2.8
20	10.3	5.3	3.0	2.6	2.5	1.9
10	7.3	3.5	1.6	1.3	1.2	.9
5	5.4	2.1	.7	.7	.6	.5
No. of pupils	327	578	702	693	473	129

TABLE VI. HARLAN'S TEST OF INFORMATION IN AMERICAN HISTORY PERCENTILE SCORES BASED ON MAY TESTING

Percentile	Grade	
	VII	VIII
95	78.8	95.2
90	70.9	94.1
80	56.6	92.2
70	54.0	81.1
60	48.8	74.9
50	43.9	68.2
40	38.8	56.7
30	31.9	50.9
20	27.1	41.3
10	20.1	31.4
5	15.5	27.7
Number of pupils	1109	1691

TABLE VII. HOTZ'S FIRST YEAR ALGEBRA SCALES, SERIES A PERCENTILE SCORES BASED ON MAY TESTING

Percentile	Addition and Subtraction		Multiplication and Division		Problems		Equations and Formulas		Graphs	
	Grade		Grade		Grade		Grade		Grade	
	IX	X	IX	X	IX	X	IX	X	IX	X
95	11.7	10.5	10.4	10.7	10.6	9.3	12.3	11.0	8.0	7.5
90	10.8	9.7	9.7	10.0	9.9	7.6	11.5	10.4	7.5	6.9
80	9.6	8.9	8.7	9.0	9.0	6.6	10.3	9.5	7.2	6.3
70	8.6	8.4	8.1	8.4	8.1	5.9	9.3	8.8	6.8	5.8
60	7.7	8.0	7.6	7.8	7.3	5.3	8.6	8.2	6.5	5.4
50	6.9	7.3	7.2	7.4	6.4	5.0	7.7	7.9	6.2	5.0
40	6.3	6.7	6.6	6.9	5.7	4.6	7.0	7.5	5.8	4.7
30	5.7	6.1	6.0	6.2	4.8	4.1	6.4	7.0	5.4	4.3
20	4.8	5.5	5.4	5.6	3.8	3.6	5.7	6.4	5.0	3.9
10	3.7	4.3	4.6	4.9	2.6	2.9	4.5	5.5	4.2	3.3
5	2.9	3.2	4.0	4.2	1.5	2.2	3.7	4.3	3.7	2.8
Number of pupils	561	390	570	388	566	394	478	385	121	413

TABLE VIII. HOLLEY'S SENTENCE VOCABULARY SCALES. PERCENTILE SCORES BASED ON APRIL TESTING

Percentile	Grade										
	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
95	72.7	53.5	45.2	50.9	65.5	63.9	66.5	68.5	75.2	77.9	82.3
90	68.5	44.7	40.0	47.6	61.0	58.5	62.6	64.6	70.0	75.0	78.2
80	63.8	32.3	35.7	42.4	54.7	52.5	57.4	58.7	66.0	69.6	73.1
70	59.6	25.7	31.6	38.5	49.2	48.2	54.2	55.5	61.8	66.4	68.8
60	57.2	20.0	28.2	35.8	46.2	45.1	50.9	52.3	58.7	63.1	65.8
50	54.8	16.7	25.1	33.0	42.8	41.9	47.8	49.0	56.1	59.9	62.8
40	52.4	13.4	22.1	30.3	39.5	38.4	44.7	45.8	53.4	56.9	59.7
30	50.0	10.1	18.7	27.0	35.5	34.3	41.6	42.6	50.8	53.9	56.3
20	46.1	6.8	14.7	23.6	31.4	30.2	36.8	38.9	46.6	51.0	52.8
10	42.2	3.4	10.7	20.2	24.9	20.9	30.0	33.9	41.6	45.4	48.3
5	40.3	1.7	6.0	13.1	21.1	13.4	22.8	31.3	37.4	42.1	44.1
No. of pupils	117	406	520	465	450	1188	1047	253	223	155	108

410.6.

THE ILLINOIS EXAMINATION

BY

Walter S. Monroe
DIRECTOR

Bureau of Educational Research
College of Education, University of Illinois

PUBLISHED BY THE UNIVERSITY OF ILLINOIS
URBANA, ILLINOIS

TABLE OF CONTENTS

PAGE

PREFACE 4

INTRODUCTION 5

I. FACTS OF TITLE 6

II. NATURE OF PUPIL'S PERFORMANCE 6
  1. Illinois General Intelligence Scale 6
  2. Monroe's Standardized Reading Tests, Revised...... 10
  3. Monroe's General Survey Scale in Arithmetic....... 12

III. DESCRIPTION OF PUPIL'S PERFORMANCE.................... 17
  1. Point Scores 17
  2. Derived Scores 19

IV. FUNCTION 43

V. VALIDITY 43
  1. Objectivity 43
  2. Reliability 43
  3. Discrimination 50
  4. Comparison with criterion measures 57
  5. Inferences concerning validity based upon the structure of the test and its administration............ 59

VI. VALIDITY OF SIGNIFICANCE 61

VII. NORMS 65

PREFACE.

In this bulletin we present an account of the derivation of the battery of educational tests known as the Illinois Examination. In addition, we include data relative to their validity, reliability, practice effect, norms, and significance. Doubtless some readers will not find all of the information relative to these tests which they desire. Although an effort has been made to have the account complete the author himself is conscious of a number of limitations of this report. It is, however, presented in hopes that it will be of service to those who are interested in the Illinois Examination.

On April 2, 1920, a "Committee on Standard Tests" of the Illinois Association of County Superintendents met in conference at the University of Illinois, in order to decide upon a testing program for rural schools. A group of three tests, including one for the measurement of general intelligence, one for measuring ability to read silently, and one for measuring abilities in the field of arithmetic, were recommended to the committee by the Bureau of Educational Research. Although the school year was far advanced it was decided to incorporate in this battery of tests, the achievement quotient which had been conceived in another connection.

It is obvious that educational research of the type represented by the bulletin is possible only through the cooperation of a large number of persons. The writer is glad to acknowledge the cooperation of, *first*, the city superintendents and teachers who gave the preliminary form of the tests in May, 1920, *second*, the county superintendents and others who cooperated in the giving of the tests in October, 1920, and *third*, those city superintendents and teachers who gave portions of the Examination in the spring of 1921 in order to determine the reliability of the tests. Acknowledgement is made in the body of the bulletin for special contributions.

WALTER S. MONROE, Director.

THE ILLINOIS EXAMINATION.

INTRODUCTION

The Illinois Examination, published in the form of a sixteen page booklet, is the name given to a battery of tests: The Illinois General Intelligence Scale, Monroe's Standardized Silent Reading Tests, Revised, and Monroe's General Survey Scale in Arithmetic.* However, the significant characteristic of the Illinois Examination is not that of merely putting these three tests together in one booklet. It is rather the way in which the measures of achievement are combined with the measure of intelligence to secure the Achievement Quotient (A. Q.) The usual procedure of interpreting measures of achievement by reference to grade norms provides for no consideration of the general intelligence of the pupil; all pupils, bright, average, and dull are judged with reference to the same norms. The plan by which the measures of achievement and general intelligence are combined in the Illinois Examination was conceived in an effort to provide a procedure by which a pupil's general intelligence or his capacity to learn would be considered in interpreting his achievement.

Briefly, the plan consists of establishing for the achievement tests mental age norms which are used to supplement the usual grade norms. For each half year of mental age, as shown by the general intelligence scale used, the median achievement was determined. These medians are the mental age norms, which are used as a basis for translating the point scores into achievement ages. In arriving at the mental age norms, all pupils of a given mental age in grades III to VIII inclusive were grouped together without regard to the grade in which they were classified. Provision is made for comparing a pupil's achievement score, (when expressed as an achievement age) with the norm corresponding to his mental age by dividing his achievement age by the standard score for his mental age.** This quotient is called the Achievement Quotient. The plan involves certain assumptions and approximations. These, as well as the validity and the reliability of the several tests incorporated in the Illinois Examination, will be considered in detail on the following pages.

In the construction of the tests crude methods were frequently employed in preference to more refined ones because of their greater simplicity and the saving of time thus secured. For example, this was done in the construction of the duplicate forms of the tests in silent reading and arithmetic and in deriving the basis for translating the point scores into age scores. In the latter case a more refined procedure involving the use of the regression equation might have been used. The results obtained by the use of the crude methods have been verified by a critical study of the test.

---

*These tests are published by the Public School Publishing Co., Bloomington, Illinois.

**This "standard score" is numerically the same as a pupil's mental age.

I. FACTS OF TITLE.

In so far as the Illinois General Intelligence Scale represents originality, B. R. Buckingham is primarily responsible for it. Walter S. Monroe, the author of this bulletin, devised the two achievement tests, and also contributed the plan of combining measures of achievement with measures of intelligence. An experimental edition of the Illinois Examination, including the scales for general intelligence and silent reading, was published in April, 1920. The complete examination together with the Teachers Handbook and class record sheet was first published during the summer of 1920 and was made available for distribution in August. The second form of the Illinois Examination was devised during the school-year of 1920-21 and was made available for use early in 1921. The Examination consists of two parts. Illinois Examination I is designed for grades III, IV and V. Illinois Examination II is designed for grades VI, VII, and VIII. The Illinois General Intelligence Scale is the same for all grades. The scales for reading and for arithmetic are different in the two examinations. There are two forms of the examination. A third form of the scales for silent reading and arithmetic has been prepared.

II. NATURE OF PUPIL'S PERFORMANCE.

1. **Illinois General Intelligence Scale.** In the Illinois General Intelligence Scale the pupil is required to make a mark with his pencil or to write figures. No other writing is required of him. The marks consist either of a line drawn under a word or of a line crossing out a word or number.

The scale consists of seven sub-tests which are illustrated by a few samples from each.

Test No. 1—ANALOGIES

1 eat—bread:: drink—**water iron lead stones**........................ 1
2 finger—hand::toe—**box foot doll coat**................................ 1
3 shoe—foot::hat—**kitten head knife penny**........................ 3
4 dress—women::feathers—**bird neck feet bill** 4
5 dog—puppy::cat—**kitten dog tiger house**........................ 5

The pupil is asked to draw a line under the word in heavy type related to the third word as the second is related to the first.

Test No. 2—ARITHMETIC PROBLEMS

1 If one boy has 10 fingers, how many fingers have six boys?............ Answer (   )
2 There are 15 children in our class. 5 of them are boys. How many are girls?.. Answer (   )
3 We learn 2 words a day in our class. How many do we learn in 8 days? Answer (   )
4 Jack is 42 inches tall and Fred is 5 inches taller. How tall is Fred? ....Answer (   )
5 Mr. Gray sold ten bags of flour last Saturday at 2 dollars a bag. How many dollars did he get for the flour? ....Answer (   )

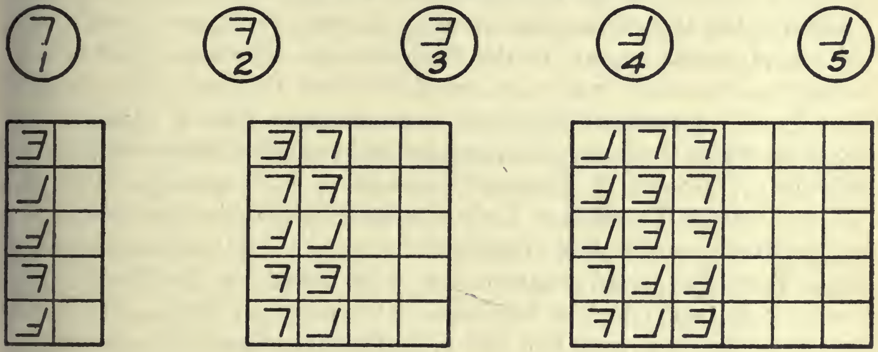
The pupil is required to write the answer within the parentheses.

Test No. 3—SENTENCE VOCABULARY

1 **A gown** is a **string animal dress plant.**
2 **Haste** is **hurry red little sweet.**
3 To **tap** is to **run fall knock smile.**
4 **A dungeon** is **open bright heavy dark.**
5 **Majesty** refers to **dresses kings countries climates.**

The pupil is required to draw a line under that one of the last four words which makes a true sentence.

Test No. 4—SUBSTITUTION



The pupil is to write in squares after the symbols the numbers to which they correspond as given by the key at the top.

Test No. 5—VERBAL INGENUITY

1 the cat at see.
2 boy was sky the sick.
3 Bread sweep will the kitchen I.
4 are going yesterday to-morrow we.
5 me mine give my straw hat.

The pupil is to cross out the superfluous word. The remaining words can then be made into a true sentence.

Test No. 6—ARITHMETICAL INGENUITY

(1) 1 2 3 9 4 5
(2) 2 4 6 7 8
(3) 9 8 7 6 5 2
(4) 11 10 8 6 4 2
(5) 5 7 10 15 20 25

The pupil is to cross out the number that does not fit in the group.

Test No. 7—SYNONYM-ANTONYM

1 high—low.......................................same—opposite 1
2 go—leave ....same—opposite 2
3 large—great ....same—opposite 3
4 bitter—sweet...................................same—opposite 4
5 begin—commence ....same—opposite 5

The pupil is required to indicate whether the first two words mean the same or the opposite by drawing a line under the word "same" or "opposite."

These sub-tests were selected after analyzing a number of existing general intelligence scales with reference to sub-tests and after examining much that has been written on the topic of group intelligence scales. The opinion of certain experts in this field was also obtained. Test No. 1, Analogies, was made from exercises taken from the Army Scale Alpha. Test No. 2, Arithmetic Problems, was compiled from a collection of problems which Professor Buckingham had evaluated with reference to difficulty. Test No. 3, Sentence Vocabulary, is an abridgment of the Holley Sentence Vocabulary Scale. In general this abridgment was secured by taking every third exercise of the original test, beginning with the first. Test No. 4, Substitution, was contributed by Professor E. H. Cameron of the College of Education, University of Illinois. Test No. 5, Verbal Ingenuity, and Test No. 6, Arithmetical Ingenuity, were taken from the Pressey Cross Out Tests.* Test No. 7, Synonym-Antonym, consists of exercises selected largely from the Army Scale Alpha.

These seven sub-tests together with two others were given to a limited number of pupils before the preliminary edition was printed. On the basis of these data two of the sub-tests were eliminated and minor changes made in the seven retained. In general the exercises in each of the sub-tests are arranged in order of gradually increasing difficulty.

---

*Acknowledgment is hereby made to Professor S. L. Pressey, of Ohio State University, for permission to use these tests.

Form 2 of the sub-tests was devised in either of two ways. The first method was to make use of "standardized exercises," either in the files of the Bureau of Educational Research, or available in published reports.* The second method involved the modification of the exercises of the first form in such a way as to change their identity with the least probable change in difficulty.

The first method was used in the case of Test 1, Analogies; Test 2, Arithmetic Problems; Test 3, Sentence Vocabulary; and Test 7, Synonym-Antonym. For Test 1, exercises published by Professor R. Pintner in the Journal of Applied Psychology, June-September, 1920, were used. For Test 2, exercises were taken from Professor Buckingham's own list of such material. The exercises for Test 3 were taken from a list of such exercises evaluated by Dr. C. E. Holley when he was engaged in the derivation of his Sentence Vocabulary Scale. The exercises for Test 7 were taken from the Army Scale Alpha as was done in the case of Form 1.

Form 2 of Test 4, Substitution, was constructed by using the same characters as in Form 1, but changing their numerical equivalents. The second form of Test 5, Verbal Ingenuity, and of Test 6, Arithmetical Ingenuity, are merely modifications of Form 1. Several modifications of each of the exercises were given to pupils and that one selected which proved to be most nearly of the same difficulty as the corresponding exercise of Form 1.

The pupil is made acquainted with the nature of the exercises in each sub-test by means of a few sample exercises in addition to a verbal explanation. With the exception of Test 2, Arithmetic Problems, three or four sample exercises are given for each sub-test. The time allowance for each of the sub-tests is given below. In each case it is intended that no pupils, even the brightest, will finish all of the exercises within the time allowed.

Test 1, Analogies,	2 minutes
Test 2, Arithmetic Problems,	3 "
Test 3, Sentence Vocabulary,	2 "
Test 4, Substitution,	3 "
Test 5, Verbal Ingenuity,	3 "
Test 6, Arithmetical Ingenuity,	2 "
Test 7, Synonym-Antonym,	1 "
Total	16 minutes

Equivalence of duplicate forms. In order to determine the degree of equivalence of the two forms of the Illinois General Intelligence Scale, copies of the two forms were arranged in alternate order and distributed to pupils as they happened to be seated. This was done in a number of school systems. The median and average scores secured from the two forms in this way are given in Table I. The median and average scores are expressed in terms of points. The differences for both the median and the average show that the forms are approximately equal. These differences vary slightly from grade to grade. The greatest difference exists in the fourth grade. Except in two instances the differences are positive which shows that the second form yields slightly larger scores than the first. This means that the second form is slightly easier than the first. When the six grades are combined the difference between the medians is 0.8. The difference of the averages is 2.3. Since, as we show later, 10.0 points of a pupil's score are equivalent to one year of mental age, we may say that in general the non-equivalence of the two forms of this scale is probably not more than two months.

---

*This standardization refers to the determination of the difficulty of the exercises upon the basis of the percent of pupils who were able to do them correctly.

TABLE I. DEGREE OF EQUIVALENCE OF FORMS 1 AND 2 OF ILLINOIS GENERAL INTELLIGENCE SCALE

Grade	No. of Pupils		Median Score		Difference	Average Score		Difference
	Form 1	Form 2	Form 1	Form 2	Form 2-Form 1	Form 1	Form 2	Form 2-Form 1
III	331	325	32.3	32.7	0.4	33.1	33.7	0.6
IV	298	295	46.0	50.5	4.5	47.3	51.5	4.2
V	336	334	56.9	56.7	−.2	58.0	59.4	1.4
VI	300	288	73.7	76.2	2.5	75.1	76.0	0.9
VII	289	279	81.9	84.1	2.2	82.6	83.8	1.2
VIII	240	253	102.1	101.3	−.8	101.3	102.7	1.4
III-VIII	1794	1774	62.5	63.3	0.8	63.8	66.1	2.3

2. **Monroe's Standardized Silent Reading Tests, Revised.** In these tests the pupil is asked to read a series of exercises which have no connection with each other. In each exercise the pupil is required to read a paragraph and to answer a question based upon it. The answer is to be given by drawing a line under a word or by indicating it in some other way. No writing is required. In most exercises the pupil has to select one out of five words in making his response; in a few there are only four words. The nature of the exercises may be illustrated by the following. The first two are taken from Test I and the others from Test II:

It was a rainy, dark, dismal day. The children had not been allowed to go out to play all day. Their lessons were poor and the teacher cross. It was late in the afternoon.

Draw a line under the word that tells how the children felt.

active smiling happy cross good

"The golden rod is yellow,
The corn is turning brown,
The trees in apple orchards
With fruit are bending down."

Draw a line under the season of the year you think is pictured in this stanza.

autumn spring winter summer

It was cold, bleak, biting weather; foggy withal; and he could hear the people in the court outside go wheezing up and down, beating their hands upon their breasts and stamping their feet upon the pavement-stones to warm them.

What kind of picture does this paragraph describe?

comfortable luxurious cheerless pleasant exciting

The caravan, stretched out upon the desert, was very picturesque; in motion, however, it was like a lazy serpent. By and by its stubborn dragging became intolerably irksome to Balthasar, patient as he was.

Place a line under the word which tells in what respect the caravan resembled a serpent.

temper color length motion size

In front the purple mountains were rising up, a distant wall. Cool snow gleamed upon the summits. Our horses suffered bitterly for water. Five hours we had ridden through all that arid waste without a pause.

What kind of a country had these people been riding through?

mountainous swampy desert forest valley

The exercises for Form 1 were taken with some modifications from the original edition of Monroe's Standardized Silent Reading Tests. In selecting the exercises, and in making the modifications, an effort was made to have all the exercises approximately the same length. Those for test II are slightly longer than the ones included in Test I. The exercises are not absolutely equal with respect to difficulty. They are arranged so that in general there is a slight increase in difficulty from exercise to exercise but in no sense can they be considered a difficulty scale. To secure absolute uniformity with respect to difficulty would have required the expenditure of a prohibitive amount of labor. Even if this were not true it is believed to be desirable to have a moderate range of difficulty in view of the fact that the exercises are to be given to pupils in a sequence of three successive school grades.

The nature of the exercises is explained to the pupils by means of three fore-exercises. In addition they are given certain verbal explanations. Four minutes are allowed for the test in all grades. This time allowance was intended to be such that practically no pupils would complete the tests; but it has been found that this is not always the case.

A few of the exercises are based on poetry although the majority are based on prose paragraphs. In light of a recent investigation* it appears likely that the reading of poetry, even for the purpose of answering a question, is not the same activity as the reading of prose for the same purpose. To the extent that this is true the test is not consistent, and this constitutes one of its limitations. In a number of other respects a high degree of uniformity has been secured. In every case the pupil's response is the same. The exercises are approximately equal in length and the questions asked appear to call for much the same type of reading.

Equivalence of duplicate forms. The three forms of each of the silent reading tests were arranged in alternate order and given in this order to pupils as they were seated in a number of cities. Table II gives the median and average comprehension scores for each of the three forms.** With few exceptions these differences are less than one. A few are negative but most are positive. In the case of Test I, when the scores from grades III to V inclusive are combined, the three forms are shown to be very nearly equivalent both by the difference of the medians and by the difference of the averages. In the case of Test II, which is designed for grades VI, VII, and VIII, Forms 1 and 2 are shown to be exactly equivalent by the medians and approximately so by the averages. Form 3, however, appears to be slightly easier, thus yielding slightly larger scores than either Form 1 or Form 2. The difference approaches a year of achievement in comprehension.***

In Table III, corresponding data are given for the rate scores. The**** differences here are larger. This is partly due to the fact that rate scores are expressed in terms of a much smaller unit than comprehension scores, but in this test the reading rate appears to be more erratic than comprehension. Although some of the differences for Test I are relatively small, corrections should be made when comparisons are made between scores obtained from the different forms.

3. **Monroe's General Survey Scale in Arithmetic.** Scale I, designed for grades III, IV, and V, consists of eight sub-tests. Scale II, designed for grades VI, VII, and VIII, consists of seven sub-tests. In each subtest, except No. 7, Scale II, the pupil is asked to do arithmetical examples. In Test 7, he is asked to insert the decimal point in quotients.

---

*Pressey, L. W. and Pressey, S. L. "A critical study of the concept of silent reading ability." *Journal of Educational Psychology*, 12:25-31, January, 1921.

**The unit in terms of which scores are expressed is relatively large. It approaches one year of achievement in comprehension of silent reading.

***Since Form 3 is not incorporated in the Illinois Examination, its non-equivalence with the other two forms does not constitute a limitation of this battery of tests.

****Except for the lower achievement ages, 7 or 8 units are equivalent to one year of achievement in rate of silent reading.

TABLE II. DEGREE OF EQUIVALENCE OF FORMS 1 AND 2 OF MONROE'S STANDARDIZED SILENT READING TESTS, REVISED (COMPREHENSION)

Grade	Number of Pupils			Median Score			Difference in Forms			Average Score			Difference in Forms		
	Form 1	Form 2	Form 3	Form 1	Form 2	Form 3	2–1	3–1	3–2	Form 1	Form 2	Form 3	2–1	3–1	3–2
III	323	299	321	5.2	4.9	5.1	–.3	–.1	0.2	5.8	5.3	5.4	–.5	–.4	0.1
IV	286	303	299	7.4	7.4	7.3	0.0	–.1	–.1	7.5	7.8	7.8	0.3	0.3	0.0
V	340	302	325	9.2	9.5	10.1	0.3	0.9	0.6	9.6	10.0	10.1	0.4	0.5	0.1
III–V	949	904	945	7.6	7.5	7.4	–.1	–.2	–.1	7.7	7.7	7.8	0.0	0.1	0.1
VI	269	272	275	11.5	11.9	11.7	0.4	0.2	–.2	11.4	12.0	12.2	0.6	0.8	0.2
VII	279	265	277	13.1	12.6	13.4	–.5	0.3	0.8	13.2	12.9	13.5	–.3	0.3	0.6
VIII	249	267	255	14.5	15.0	16.5	0.5	2.0	1.5	14.5	15.1	16.4	0.6	1.9	1.3
VI–VIII	797	804	807	13.0	13.0	13.6	0.0	0.6	0.6	13.0	13.3	14.0	0.3	1.0	0.7

TABLE III. DEGREE OF EQUIVALENCE OF FORMS 1 AND 2 OF MONROE'S STANDARDIZED SILENT READING TESTS, REVISED (RATE)

Grade	Number of Pupils			Median Score			Difference in Forms			Average Score			Difference in Forms		
	Form 1	Form 2	Form 3	Form 1	Form 2	Form 3	2–1	3–1	3–2	Form 1	Form 2	Form 3	2–1	3–1	3–2
III	323	299	321	86.9	82.3	85.6	–4.6	–1.3	3.3	90.2	83.8	93.1	–6.4	2.9	9.3
IV	286	303	299	113.7	109.0	110.1	–4.7	–3.6	1.1	113.6	118.2	117.2	4.6	3.6	–1.0
V	340	302	325	138.7	129.1	143.6	–9.6	4.9	14.5	133.4	134.8	141.0	1.4	7.6	6.2
III-V	949	904	945	116.9	104.7	110.0	-12.2	–6.9	5.3	112.7	112.3	117.2	–.4	4.5	4.9
VI	269	272	275	155.2	160.7	168.8	5.5	13.6	8.1	153.9	161.5	170.5	7.6	16.6	9.0
VII	279	265	277	166.5	167.4	181.0	0.9	14.5	13.6	172.3	169.0	183.0	–3.3	10.7	14.0
VIII	249	267	255	187.2	189.8	223.2	2.6	36.0	33.4	187.4	196.8	225.5	9.4	38.1	28.7
VI–VIII	797	804	807	165.7	173.0	183.5	7.3	17.8	10.5	170.8	175.6	192.2	4.8	21.4	16.6

In selecting the sub-tests for each scale, an effort was made to include examples of the types most appropriate for the pupils to whom they would be given. Tests 1, 2, 3, and 4 of Scale I are on the fundamental combinations, or tables, one test being devoted to each operation. These tests are similar to the corresponding tests of the Courtis Standard Research Tests, Series A, and of the Cleveland Survey Tests. Test 5 calls for single column addition of five figures. Test 6 consists of subtraction examples, in which the subtrahend is a single figure. Test 7, multiplication, consists of multiplication examples in which the multiplier is a single figure. In Test 8, division, the divisor is a single figure.

The sub-tests of Scale II are represented by the following samples:

Test No. 1—ADDITION

7862	6809	8941	5917	6772	7864	1249
5013	7623	7910	4814	6028	7883	8975
1761	5299	9845	9007	6535	8240	9005
5872	6601	8522	6975	2340	9869	1573
3739	3496	1046	1227	2319	6794	3203

Test No. 2—MULTIPLICATION

4857	5718	6942	4065
36	92	58	47

Test No. 3—DIVISION

41)574	79)36893	32)384	58)27608

Test No. 4—SUBTRACTION

739	1852	975	1087	516	962
367	948	906	821	239	325

Test No. 5—ADDITION AND SUBTRACTION OF FRACTIONS

$\frac{1}{6}+\frac{1}{3}=$ $\frac{3}{4}-\frac{2}{5}=$ $\frac{1}{6}+\frac{3}{5}=$

Test No. 6—MULTIPLICATION AND DIVISION OF FRACTIONS

$\frac{2}{3}\times\frac{3}{4}=$ $\frac{4}{7}\div\frac{2}{3}=$ $\frac{5}{12}\times\frac{3}{5}=$

Test No. 7—DECIMAL FRACTIONS

.03)16.2	Ans. :54	.07)1.82	Ans.: 26	.05).415	Ans.: 83
.06)7.44	Ans.: 124	.08).952	Ans.: 119	.04)87.6	Ans.: 219
.02).144	Ans.: 72	.08)40.8	Ans.: 51	.09)3.42	Ans.: 38

Several of the sub-tests were taken bodily from Monroe's Diagnostic Tests in Arithmetic. Others were constructed similar to well known tests. In constructing the duplicate forms of these tests the figures of the examples were rearranged or changed, so that examples identical in gross structure but leading to different answers were obtained. In the case of Scale I, Tests 1, 2, 3, and 4 were constructed by securing at random numbers of the same general magnitude as those used in Form 1. In doing this, especially in Tests 1 and 3, many of the examples were repeated either in their identical form or with the position of the two numbers reversed. Test 5 was constructed by rearranging the same numbers actually used in Test 5 of Form 1. This rearrangement was largely, although not entirely so, a process of shifting the position of the example and reversing the order of numbers therein. In Test 6, two methods were used, either the subtrahend and minuend were grouped together differently or the figures in the minuend were reversed. In Test 7, the figures in the multiplicand were rearranged and the resultant number grouped with another multiplier. In Test 8, the same divisors were kept and the dividends either slightly increased or decreased so as to leave the quotients still integral numbers.

In Test 1 of Scale II, the columns of figures in each example were shifted with occasional changes to prevent a zero from coming first in any number. In Test 2, the figures of the multiplicand were rearranged and the resultant number grouped with another multiplier. In Test 3, either the dividends were slightly increased or decreased so as to be evenly divisible by the same divisor, or the dividend or divisor was multiplied or divided by two. In Test 4, either the figures of the minuend were arranged differently or new minuends were constructed by grouping together part of the figures from two minuends to form one, the remaining figures to form another. The subtrahends were either changed by a different arrangement of figures or by adding or subtracting a small number, such as ten or twenty, or they were not changed at all. Tests 5 and 6 were constructed by a random selection of fractions of the same general magnitude as those of the same test in Form 1. In Test 7, the position of the decimal point in either divisor or dividend or in both was shifted, and the position of the examples was also changed.

In almost all of the tests an occasional change, not covered by the statements above, was made, either because of the necessity of avoiding impossible combinations or because the result by following too closely the procedure laid down seemed undesirable.

Equivalence of Duplicate Forms. A study was made of the equivalence of the duplicate forms of Monroe's General Survey Scale in Arithmetic in the same way that the equivalence of the duplicate forms of tests for general intelligence and silent reading was investigated. The results are given in Table IV. Since Forms 2 and 3 were constructed from Form 1 by a rather mechanical procedure one might expect to find a higher degree of equivalence for the arithmetic scale than for the others. This, especially in the upper grades, is not true. In the case of Scale I, however, the differences are relatively small, particularly as shown by the averages.* In the case of Scale II the differences between Form 1 and Form 2 are small. Hence, these two forms may be considered as being very approximately equivalent. Form 3 appears to be considerably easier than the other two.

III. DESCRIPTION OF PUPIL'S PERFORMANCE.

Point Scores. The performance of a pupil upon each of the sub-tests of the Illinois General Intelligence Scale is described in terms of a point score. With two exceptions this point score is the number of exercises which he has done correctly in the time allowed. In order to give appropriate weight to Test No. 4, Substitution, the number of figures written correctly is divided by four. Since the possible score on this test is 150, as compared with 16 to 40 in the cases of the other tests, failure to weight the score from this test would give it an exceedingly high degree of potency in determining the pupil's total score. The rule to divide this score by 4 is purely arbitrary. In Test No. 7, Synonym-Antonym, in order to discount for the correct answers which might be obtained by merely guessing, the pupil's score is given as the number of exercises right minus the number wrong; exercises skipped are counted as wrong; exercises not reached when time is called are not counted. The pupil's total point score is the sum of his scores on the seven sub-tests.

In the case of Monroe's Standardized Silent Reading Tests, Revised, the pupil receives two scores. His comprehension score is the number of exercises which he answers correctly. His rate score is the number of words which he reads per minute or the total number of words read divided by 4. In order to obtain this rate the pupil is asked to mark the line which he is reading when the signal to stop is given. The cumulative totals of the words are printed in the left hand margin so that the total number of words read is easily obtained.

The number of examples right is taken as the pupil's point score on each of the sub-tests of Monroe's General Survey Scale in Arithmetic. These sub-tests yield scores which differ widely in magnitude. Arbitrary

---

*The equivalence of one year of achievement is 8 units of the point score.

TABLE IV. DEGREE OF EQUIVALENCE OF FORMS 1 AND 2 OF MONROE'S GENERAL SURVEY SCALE IN ARITHMETIC

Grade	Number of Pupils			Median Score			Difference in Forms			Average Score			Difference in Forms		
	Form 1	Form 2	Form 3	Form 1	Form 2	Form 3	2–1	3–1	3–2	Form 1	Form 2	Form 3	2–1	3–1	3–2
III	225	222	227	19.6	20.6	20.5	1.0	0.9	–.1	20.9	21.2	22.1	0.3	1.2	0.9
IV	205	207	188	33.1	37.3	33.9	4.2	0.8	–3.4	38.1	40.1	38.5	2.0	0.4	–1.6
V	219	223	219	48.4	50.8	47.2	2.4	–1.2	–3.6	50.5	51.0	49.4	0.5	–1.1	–1.6
III–V	649	652	634	32.8	34.8	33.7	2.0	0.9	–1.1	33.1	34.0	33.0	0.9	–.1	–1.0
VI	222	217	210	53.6	56.5	62.6	2.9	9.0	6.1	55.5	58.8	66.4	3.3	10.9	7.6
VII	204	203	201	61.5	59.9	68.9	–1.6	7.4	9.0	63.9	62.3	70.1	–1.6	6.2	7.8
VIII	192	202	198	78.0	80.0	75.7	2.0	–2.3	–4.3	82.1	81.6	76.0	–.5	–6.1	–5.6
VI–VIII	618	622	609	64.1	64.2	68.3	0.1	4.2	4.1	66.5	67.0	70.6	0.5	4.1	3.6

rules were adopted for weighting these scores so that approximately equal weight would be given to each test. (For details see the tests.) The weighted sum of the scores on the several sub-tests is the pupil's point score.

Correction of scores derived from the scales designed for grades VI, VII, and VIII. In the cases of both silent reading and arithmetic, the tests for grades III, IV, and V are entirely different from those given in grades VI, VII, and VIII. This makes the scores from the two sequences of grades incomparable. They have a different zero point and it is not unreasonable to expect that they would be expressed in terms of a different unit. It is relatively easy to estimate the approximate difference between the zero points. This difference can be used as a correction to be added to the point scores obtained from the tests for the upper sequence of grades. This was done and it appears that the differences in the units are not sufficiently large to introduce serious inaccuracies when the corrected scores are considered comparable.

The method of estimating the difference between the zero points may be illustrated by the case of silent reading. The medians of the point scores without correction were calculated for each grade. These were then represented graphically as shown in Figure 1. The curve of progress for grades III, IV, and V was then extended so that the extension would parallel the curve of progress for grades VI, VII, and VIII. This extension together with the progress curve for the lower sequence of grades forms a progress curve for grades III to VIII inclusive. The distance between the extension and the original curve is assumed to represent the difference in the zero points of the two tests.

The estimated corrections based upon the scores derived from the preliminary tests in May, 1920, for Monroe's Standardized Silent Reading Test II, Revised, were rate 17 and comprehension 4. The arithmetic tests were not given at this time but, on the basis of scores derived from similar tests, a correction of 16 was estimated as the proper amount to add to the point scores derived from Scale II. The scores derived from giving the Illinois Examination to fifty thousand children during the fall of 1920 indicated that these estimates were incorrect. The revised estimates are: reading rate 29, reading comprehension 5, and arithmetic 22. When these corrections are added to the point scores derived from the corresponding tests in Examination II, the scores will, in general, be approximately comparable to the corresponding scores derived from Examination I.

2. **Derived scores.** The point scores yielded by the scales of the Illinois Examination are translated into age scores. From these quotients

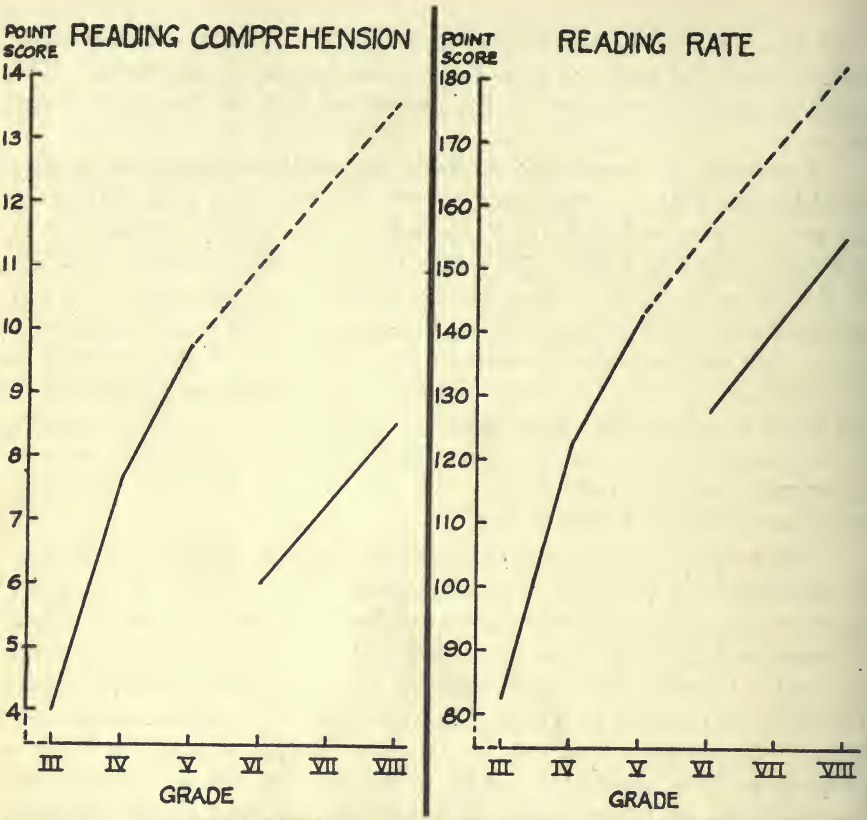


FIGURE 1. SHOWING METHOD OF ESTIMATING CORRECTIONS TO BE ADDED TO SCORES OF SCALE II.

are calculated. The portion of the Illinois Examination devoted to the measurement of general intelligence and ability in silent reading was given to over seventy-five hundred pupils in May, 1920, for the purpose of standardization and of determination of the basis for calculating certain derived scores. When the complete examination was used in October, 1920, the scores obtained indicated that the results secured in May involved a sufficiently large error to make necessary a re-determination of the grade norms and of the basis for calculating the derived scores. Unfortunately, the data collected from the October testing were in such form that the determination of the basis for calculating the derived scores could not be worked out by direct methods. It was, therefore, necessary to use estimates instead of actual determinations, but these estimates have been verified whenever possible and appear to be approximately correct.

Data, in the form required for the direct calculation of the basis for obtaining the derived scores, are available from the public schools o

Decatur, Illinois, and from eight elementary schools of the city of Chicago. These data, particularly those from Decatur, do not appear to be entirely representative. However, it has seemed wise to describe the method in terms of the scores from these two cities. The estimates will be found to agree, in some instances, with the data collected in these two cities, but in others agreement will be lacking.

The chronological age norms for the Illinois General Intelligence Scale are used as a basis for translating the point scores into mental age scores. In determining the age norms the distribution of point scores was secured for each half year of chronological age. For example, the point scores of all pupils whose chronological ages fell between nine years, and nine years and six months, were brought together in a distribution. The median of this distribution was taken as the age norm for this chronological age group. In Table V, the median point scores for each of the age groups are given for both Decatur and Chicago. In the second column of the table the average chronological age of the group is given. It will be noted that except for the extreme ages the median point scores tend to become larger as the pupils become older.

This scale was given only to the pupils in grades III to VIII inclusive. If a pupil enters school at six years we may expect to find him in the third grade at eight years of age. Of course, if the pupil fails of promotion, he will not reach the third grade until he is older. Pupils younger than eight years attaining the third grade either have entered school before they were six or have skipped a grade. In this case the pupils are probably bright. The same thing is true to a somewhat less degree for the pupils whose ages fall in the 8.0-8.4 interval, and even in the interval from 8.5-8.9 there is some selection of the brighter pupils. All pupils eight years of age have not advanced to the third grade. Some of them have failed of promotion and are found in the first or second grade. In general, they are the duller pupils. Consequently, we may say that the age groups below nine years are selected so that they do not include all pupils whose ages fall within the groups. The less capable have been left out. In Chicago, the Illinois General Intelligence Scale was not given to III-B pupils. For this reason, the selection extends above nine years.

A pupil who entered school at six years of age and was promoted each year would be found at fourteen in the ninth grade or first year of the high school. Furthermore, when pupils reach the age of fourteen, the compulsory attendance law does not apply with the same force as in the case of younger pupils. For these reasons the age-groups at the top of the table are selected. In general, they include pupils who have failed of promotion one or more times because they were not able to do the work of the school satisfactorily. Hence, we should expect to find the median point scores for such pupils less than those for younger groups of unselected pupils.

TABLE V. MEDIAN POINT SCORES OF CHRONOLOGICAL AGE GROUPS FOR DECATUR AND FOR EIGHT ELEMENTARY SCHOOLS IN CHICAGO

Chronological Age Interval	Average Chronological Age	Decatur		Chicago	
		No. of Pupils	Median Point Score	No. of Pupils	Median Point Score
17.5–17.9	17.75	1	40.0		
17.0–17.4	17.25				
16.5–16.9	16.75	2	54.5	5	57.5
16.0–16.4	16.25	9	70.0	11	61.7
15.5–15.9	15.75	26	64.5	37	67.9
15.0–15.4	15.25	61	64.4	78	75.3
14.5–14.9	14.75	124	71.4	130	72.4
14.0–14.4	14.25	191	68.9	284	70.0
13.5–13.9	13.75	244	73.5	366	67.6
13.0–13.4	13.25	297	70.2	408	72.2
12.5–12.9	12.75	306	67.2	326	67.1
12.0–12.4	12.25	304	63.4	354	63.1
11.5–11.9	11.75	308	52.8	294	57.8
11.0–11.4	11.25	298	54.8	328	55.2
10.5–10.9	10.75	316	43.8	244	46.9
10.0–10.4	10.25	290	40.9	313	48.0
9.5–9.9	9.75	297	35.9	205	44.4
9.0–9.4	9.25	285	30.1	168	43.2
8.5–8.9	8.75	216	26.2	78	40.0
8.0–8.4	8.25	163	20.0	24	38.3
7.5–7.9	7.75	33	27.5	3	42.5
7.0–7.4	7.25	14	23.0		
6.5–6.9	6.75	2	20.0		

In Figure 2, the median point scores given in Table V are represented graphically. The Decatur scores are represented by small crosses and the Chicago scores by small circles. With the exception of the scores for the older pupils there is a suggestion of a straight line relationship between the point scores and chronological age. Other considerations, some of which will be mentioned later, led to the adoption of a straight line as representing the relation between the median point scores and the chronological ages. This line has been drawn in the figure. It is considered to extend indefinitely upward. It is extended downward until it cuts the vertical axis. This straight-line relationship permits the formulation of a very simple rule for translating point scores into mental ages.

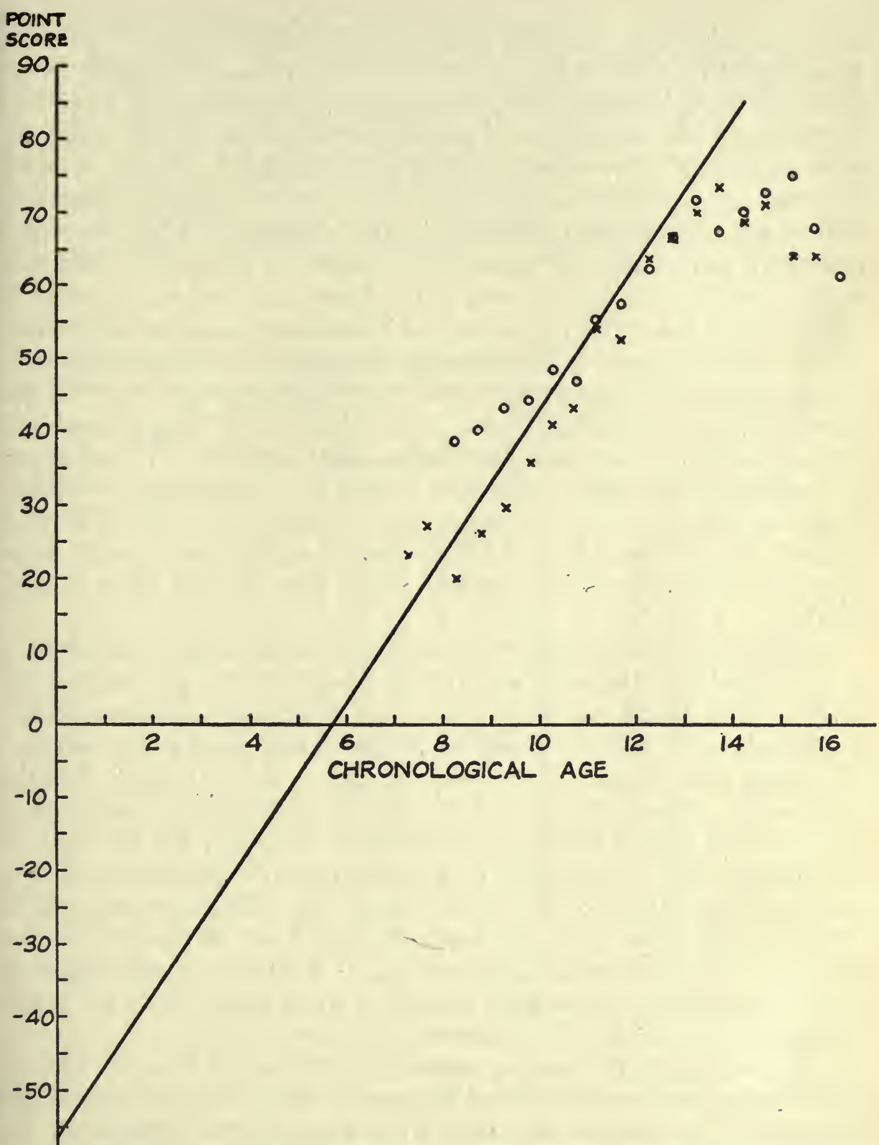


FIGURE 2. RELATION BETWEEN MEDIAN POINT SCORES AND CHRONOLOGICAL AGE. DECATUR MEDIAN SCORES REPRESENTED BY CROSSES AND CHICAGO MEDIAN SCORES BY SMALL CIRCLES.

With reference to the extension of the line of relationship upward, it appears that the development of general intelligence, like physical development, does not continue indefinitely. The evidence at hand indicates that beginning about 14 years there is a slowing down of development and that at about the age of 16 or 18 it ceases altogether* After this age, mental changes consist primarily in increasing the scope and skillfulness of the application of intelligence which one already possesses. An extension of the line of relationship which might be judged to conform with the actual median point scores for the higher age groups (assuming that all pupils belonging to these age groups were tested) would be curved downward and would tend to become horizontal or nearly so. This procedure, which is the one followed in the Binet Scale and in a number of others, would not give us a method for translating the point scores made by the bright children in the upper grades into the corresponding ages. For example, a few pupils have made point scores as high as 185 on the Illinois General Intelligence Scale. On the basis of the facts at hand it is impossible to conceive that the median point score for any age group could possibly equal this. Therefore, if such point scores are to be translated into mental ages it is necessary to devise another plan. The plan adopted was to extend the straight line of relationship upward until maximum point scores were reached. This plan has the merit of simplicity.

If the downward extension of the line of relationship is examined it will be found that the point score which is equivalent to zero mental age is -55. This means that the absolute zero of the Illinois General Intelligence Scale is -55 points. Therefore, in translating point scores into corresponding mental ages, it is necessary to correct this by adding 55 to the point score. When the slope of the line of relationship is examined it will be noted that the rise is approximately 10 points for one year of chronological age, which means that 10 points on the intelligence scale is considered equivalent to one year of mental age. Hence, we are able to state the rule, "Add 55 to the total point score and divide by 10. The quotient is the corresponding mental age." This rule is not absolutely accurate but it does not appear to involve errors greater than the errors of measurement which will be discussed in a later section.

The validity of the rule for determining the mental ages was verified by comparing the results obtained by using it with the mental ages as determined by the Stanford Revision of the Binet Scale for Measuring Intelligence. Table VI gives the determinations of mental age by the two scales for 201 pupils.** Because of the small number of pupils in some groups it is to be expected that some of the differences in mental age would be relatively large. An examination of the differences given in the last column of the table reveals that only three out of twenty are greater than one year. Many of the differences are very small. This table indicates that the rule adopted for translating point scores into mental ages gives results agreeing, as closely as could be expected in view of the errors of measurement which occur in using both scales, with the results obtained by using the Stanford Revision of the Binet Scale for Measuring Intelligence.

---

*Children probably differ with reference to the development of their general intelligence. This slowing down probably begins at different ages in the cases of different pupils. Thus this statement should be interpreted as representing the average or typical development.

**A portion of these data was furnished the writer by Professor J. C. DeVoss, of the Kansas State Normal School, Emporia, Kansas. The remainder was contributed by Superintendent L. W. Keeler, Michigan City, Indiana.

TABLE VI. COMPARISON OF MENTAL AGES OF 201 PUPILS BY BINET SCALE AND BY ILLINOIS GENERAL INTELLIGENCE SCALE

Number of Pupils	Binet Scale		Illinois Scale		Difference
	Mental Age Interval	Mental Age Average	Med. Point Score	Corresp'd'g Mental Age	in Mental Ages
2	17.5–17.9	17.75	125.0	18.00	–.25
	17.0–17.4	17.25			
2	16.5–16.9	16.75	115.0	17.00	–.25
2	16.0–16.4	16.25	130.0	18.50	–2.25
8	15.5–15.9	15.75	90.0	14.50	1.25
6	15.0–15.4	15.25	105.0	16.00	–.75
5	14.5–14.9	14.75	75.0	13.00	1.75
9	14.0–14.4	14.25	85.0	14.00	.25
13	13.5–13.9	13.75	92.1	14.71	–.96
19	13.0–13.4	13.25	78.8	13.38	–.13
11	12.5–12.9	12.75	71.7	12.67	.08
11	12.0–12.4	12.25	63.0	11.80	.45
18	11.5–11.9	11.75	64.0	11.90	–.15
32	11.0–11.4	11.25	58.8	11.38	–.13
16	10.5–10.9	10.75	57.5	11.25	–.50
18	10.0–10.4	10.25	44.0	9.90	.35
10	9.5–9.9	9.75	50.0	10.50	–.75
8	9.0–9.4	9.25	36.3	9.13	.12
5	8.5–8.9	8.75	28.3	8.33	.42
3	8.0–8.4	8.25	27.5	8.25	.00
3	7.5–7.9	7.75	25.0	8.00	–.25

The original data* from which Table VI was obtained are represented graphically in Figure 3. The coordinates of each dot in this figure represent the mental ages of a pupil as determined by the two scales. The abscissa represents mental age as determined by the Illinois Examination. The ordinate represents the mental age as determined by the Binet Scale. The two instruments yield approximately the same mental ages for a number of pupils. For others the difference is marked. It must be remembered that neither of the instruments is perfect and that when a difference exists between the two mental ages it is likely to be due to the fact that both rather than only the mental age yielded by the Illinois General Intelligence Scale involve errors.

---

*Two additional cases are represented in the figure.

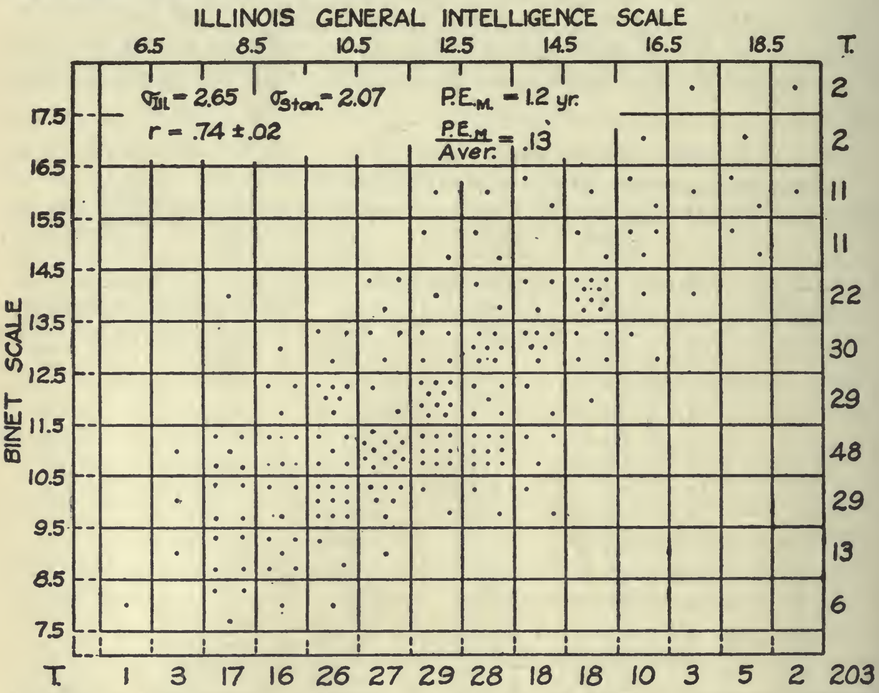


FIGURE 3. RELATION OF MENTAL AGES OF 203 PUPILS AS DETERMINED BY THE ILLINOIS GENERAL INTELLIGENCE SCALE AND BY THE BINET SCALE.

It should be borne in mind that mental ages above 14 obtained by the Illinois General Intelligence Scale must be interpreted somewhat differently from the mental ages obtained by other scales, especially the Stanford Revision of the Binet Scale. The extension of the concept of mental age which is here introduced is believed to be justified by the practical consideration of finding a convenient device to represent the absolute intelligence of bright children in the upper grades.

The intelligence quotient (I. Q.) is a measure derived by dividing a pupil's mental age by his chronological age. More strictly speaking the I. Q. is the quotient of a pupil's mental age divided by the median mental age for his chronological age. Below the age of 14 these are identical. A pupil's I. Q. is an index of his intelligence. If it is 100 he possesses only average intelligence; if it is greater than 100, he is brighter than the average; if less than 100 he is duller. Because the mental ages above 14 do not have the same meaning as the corresponding mental ages obtained from the use of the Stanford Revision of the Binet Scale, it follows that the I. Q.'s for these ages also have a modified significance. For this reason reason Otis has called the quotient which we are using the "coefficient of brightness." Since below the age of 14 the quotient yielded by the Illinois General Intelligence Scale is identical in meaning with the quotient derived by the Stanford Revision of the Binet Scale, it has seemed wise to retain the name, intelligence quotient.

To facilitate the calculation of the I. Q. a table has been prepared which gives the I. Q.'s for each half year. This is reproduced as Table VII. The quotients have been calculated with reference to the mid-point of the intervals. Thus the quotient for chronological age group, 12-6, and mental age, 16-0, was found by dividing 16.25 by 12.75. A slightly different procedure was followed for chronological ages between 14 and 18. As noted above it is assumed that within this period there is a gradual slowing down of the growth of general intelligence. Hence, the differences between the actual median point scores for successive chronological age groups will gradually decrease and become zero at 18. Although the intelligence quotient is generally defined as the quotient obtained by dividing a pupil's mental age by his chronological age it is essentially the quotient obtained by dividing his mental age by the mental age norm for his chronological age. Below the age of 14 the norm and the pupil's chronological age are numerically the same. Above the age of 14 they are different. The quotients in Table VII for the age of 14 and above were calculated by using the mental age norm instead of the chronological age as the divisor.

The first column of the table contains point scores corresponding to the mental ages (M. A.'s) given in the second column. For example, a point score of 60 corresponds to a mental age of 11 years and six months, and a point score of 65 is equivalent to a mental age of 12 years. If a pupil's point score is 61, 62, 63, or 64, it is taken, when using this table, as meaning a mental age of 11 years and six months. Similarly, a pupil's chronological age is considered to be 10 years until it is 10 years and six months.

TABLE VII–INTELLIGENCE QUOTIENTS*

Point Score	Mental Age	Chronological Age																										
		5–0	5–6	6–0	6–6	7–0	7–6	8–0	8–6	9–0	9–6	10–0	10–6	11–0	11–6	12–0	12–6	13–0	13–6	14–0	14–6	15–0	15–6	16–0	16–6	17–0	17–6	18–0
185	24–0																			170	166	163	161	158	157	156	155	154
180	23–6																		173	167	162	160	157	155	154	152	152	151
175	23–0																		169	163	159	156	154	152	150	149	148	147
170	22–6																	172	165	160	155	153	151	149	147	146	145	144
165	22–0																175	168	162	156	152	150	147	145	144	143	142	141
160	21–6																171	164	158	153	148	146	144	142	141	140	139	138
155	21–0															173	167	160	155	149	145	143	140	139	137	136	135	135
150	20–6															169	163	157	151	146	141	139	137	135	134	133	132	131
145	20–0													180	172	165	159	153	147	142	138	136	134	132	131	130	129	128
140	19–6													176	168	161	155	149	144	139	134	133	130	129	127	127	126	125
135	19–0												179	171	164	157	151	145	140	135	131	129	127	125	124	124	123	122
130	18–6												174	167	160	153	147	142	136	132	128	126	124	122	121	120	119	119
125	18–0											178	170	162	155	149	143	138	133	128	124	122	120	119	118	117	116	115
120	17–6											173	165	158	151	145	139	134	129	125	121	119	117	115	114	114	113	112
115	17–0										177	168	160	153	147	141	135	130	125	121	117	116	114	112	111	111	110	109
110	16–6									181	172	163	156	149	143	137	131	126	122	118	114	112	110	109	108	107	106	106
105	16–0									176	167	159	151	144	138	133	127	123	118	114	110	109	107	106	105	104	103	103
100	15–6								180	170	162	154	147	140	134	129	124	119	115	111	107	105	104	102	101	101	100	99
95	15–0								174	165	156	149	142	136	130	124	120	115	111	107	103	102	100	99	98	97	97	96
90	14–6							179	169	159	151	144	137	131	126	120	116	111	107	104	100	99	97	96	95	94	94	93
85	14–0						184	173	163	154	146	139	133	127	121	116	112	108	104	100	97	95	94	92	92	91	90	90
80	13–6						177	167	157	149	141	134	128	122	117	112	108	104	100	96	93	92	90	89	88	88	87	87
75	13–0					183	171	161	151	143	136	129	123	118	113	108	104	100	96	93	90	88	87	86	85	84	84	83
70	12–6					176	165	155	146	138	131	124	119	113	109	104	100	96	93	89	86	85	84	83	82	81	81	80
65	12–0				181	169	158	148	140	132	126	120	114	109	104	100	96	92	89	86	83	82	80	79	78	78	77	77
60	11–6				174	162	152	142	134	127	121	115	109	104	100	96	92	89	85	82	79	78	77	76	75	75	74	74
55	11–0			180	167	155	145	136	129	122	115	110	105	100	96	92	88	85	82	79	76	75	74	73	72	72	71	71
50	10–6			172	159	148	139	130	123	116	110	105	100	96	91	88	84	81	78	75	72	71	70	69	68	68	68	67
45	10–0		178	164	152	141	132	124	117	111	105	100	95	91	87	84	80	77	75	72	69	68	67	66	65	65	65	64
40	9–6	186	170	156	144	134	126	118	111	105	100	95	91	87	83	80	76	74	71	68	66	65	64	63	62	62	61	61
35	9–0	176	161	148	137	127	119	112	106	100	95	90	86	82	79	76	73	70	67	65	62	61	60	59	59	58	58	58
30	8–6	167	152	140	130	120	113	106	100	95	90	85	81	78	74	71	69	66	64	61	59	58	57	56	56	55	55	54
25	8–0	157	143	132	122	114	106	100	94	89	85	80	77	73	70	67	65	62	60	58	55	54	54	53	52	52	52	51
20	7–6	148	134	124	115	107	100	94	89	84	79	76	72	69	66	63	61	58	56	54	52	51	50	50	49	49	48	48
15	7–0	138	126	116	107	100	94	88	83	78	74	71	67	64	62	59	57	55	53	51	48							
10	6–6	129	117	108	100	93	87	82	77	73	69	66	63	60	57	55	53	51	49	47								
5	6–0	119	109	100	93	86	81	76	71	68	64	61	58	56	53	51	49	47	45	44								
0	5–6	110	100	92	85	79	74	70	66	62	59	56	53	51	49	47	45	43	42	40								
		5–0	5–6	6–0	6–6	7–0	7–6	8–0	8–6	9–0	9–6	10–0	10–6	11–0	11–6	12–0	12–6	13–0	13–6	14–0	14–6	15–0	15–6	16–0	16–6	17–0	17–6	18–0

Chronological Age

To find a pupil's I. Q. in this table proceed as follows: Take, for example, a pupil whose point score is 82 and whose chronological age is 11 years and four months. Find the interval* in the first column of numbers which contains a point score of 82. This is the 80 interval and the corresponding mental age (M. A.) is 13 years and 6 months. Locate at the top of the table the interval for a chronological age of 11 years; follow down the column for this interval to the line for the point score of 80. The number at the intersection of this column and this line is the pupil's I. Q. It is 122. *If a pupil's chronological age is more than 18 years, call it 18.* His I. Q. will always be found in the last column of the table.

As we shall show later there is evidence that the intelligence quotients yielded by this test have a greater variability than those obtained from the Stanford Revision of the Binet Scale. Therefore, when interpreting the I. Q.'s derived from the Illinois General Intelligence Scale in terms of degrees of brightness, it will be necessary to use a basis different from that proposed by Terman and others. (See page 67).

The distributions of the intelligence quotients furnish additional evidence of the validity of the rule for translating the point scores into mental ages. In Figure 4, the distribution of the I. Q.'s of the pupils of Decatur is represented graphically. In forming this distribution all grades have been combined, but the shape of the distributions for the separate half grades approximates that of this total distribution. The total number of pupils is 3787. The only conspicuous departure from the normal curve is in the

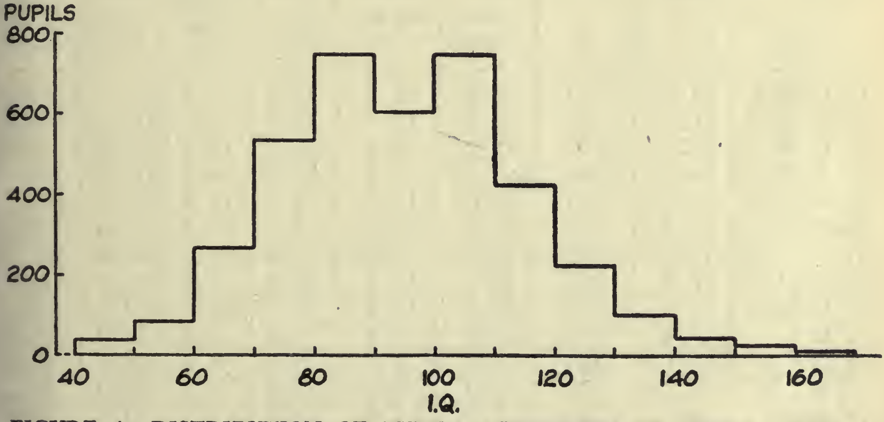


FIGURE 4. DISTRIBUTION OF PUPILS ACCORDING TO THEIR INTELLIGENCE QUOTIENTS.

---

*The term "interval" is used to indicate a span of point scores or ages. For the intelligence scores the span is five scores and in the table only the lower limit of this interval or span is given. For ages the interval is six months, and again only the lower limit is given.

90-100 interval. This is due to the structure of the table used for calculating I. Q.'s. Thus, we may say that the distributions of I. Q.'s do not indicate any error resulting from the rule used for translating point scores into mental ages.

Derived scores for achievement tests. To facilitate the comparison of a pupil's achievement with his mental age which represents his capacity to learn, provision was made for translating the point scores derived from the tests on arithmetic and silent reading into achievement age scores. The method of arriving at the basis for making this translation is similar to that used in the case of an intelligence scale. The point scores derived from all grades were thrown together and a distribution of the achievement scores was secured for each mental age group.* Because the data collected in May, 1920, were not representative, it is again necessary to illustrate the procedure and then to give the estimates which are recommended for use. Table VIII gives the median achievement point scores for the mental age groups for the city of Decatur. In this table the interval of mental

TABLE VIII. MEDIAN ACHIEVEMENT POINT SCORES FOR MENTAL AGE GROUPS IN DECATUR

Mental Age Interval	Average Mental Age	Number of Pupils	Arithmetic		Silent Reading Comprehension		Silent Reading Rate	
			Median Point Score	Corresponding Age Score	Median Point Score	Corresponding Age Score	Median Point Score	Corresponding Age Score
18.0–18.9	18.5	13	78.4	16.3	17.7	18.2	233.2	21.2
17.0–17.9	17.5	26	57.6	13.7	15.5	16.0	206.0	17.5
16.0–16.9	16.5	62	63.2	14.4	15.5	16.0	218.0	19.0
15.0–15.9	15.5	146	62.4	14.3	14.4	14.9	202.8	17.1
14.0–14.9	14.5	211	56.0	13.5	13.9	14.4	185.2	14.9
13.0–13.9	13.5	323	53.2	12.9	13.0	13.5	170.0	13.0
12.0–12.9	12.5	432	48.0	12.5	12.0	13.0	157.6	11.8
11.0–11.9	11.5	485	40.8	11.7	10.0	11.0	148.0	11.0
10.0–10.9	10.5	481	34.4	10.8	9.7	10.7	139.4	10.4
9.0–9.9	9.5	479	24.0	9.5	8.2	9.6	124.0	9.5
8.0–8.9	8.5	487	16.0	8.6	6.6	8.8	105.2	8.7
7.0–7.9	7.5	373	10.0	8.0	4.4	8.1	88.0	8.1
6.0–6.9	6.5	251	6.0	7.6	3.0	7.5	70.6	7.6
5.0–5.9	5.5	23	5.0	7.5	.5	6.9	36.0	6.9

---

*Before doing this the achievement point scores derived from the scales designed for the upper sequence of grades were corrected by adding the numbers given on page 19.

age is one year.* There are 432 pupils whose mental ages fall between 12 and 13 years. Their average mental age is 12.5. Their median point score on the arithmetic scale is 48.0. This corresponds to an achievement age of 12.5. That is, the median achievement of the pupils belonging to a mental age group is taken as the equivalent of the corresponding achievement age. In comprehension of silent reading the median point score of these pupils was 12.0. This is taken as corresponding to the achievement age of 13.0.** In rate of silent reading their median point score was 157.6. This is taken as corresponding to an achievement age of 11.8. The lack of agreement between the achievement age in rate of silent reading and the average mental age of this group is due to the fact that this group has not made as high a score in rate of reading as pupils of this mental age normally do. It must be remembered that these tables are based upon the scores of one city only. In addition to the data secured from Decatur, it was possible to make use of the grade norms derived for the Illinois Examination from 49,500 scores. Table IX gives the median point scores and the median mental age for each grade.

TABLE IX. GRADE NORMS FOR ILLINOIS EXAMINATION

Grade	Median Mental Age	Arithmetic	Silent Reading	
			Comprehension	Rate
III	7.9	10	3.8	82
IV	9.4	21	7.7	122
V	10.7	35	9.8	142
VI	12.0	44	11.0	158
VII	13.1	53	12.1	170
VIII	14.3	60	13.5	183

*A mental age interval of six months is preferable. It is not used in this illustration because the data were originally tabulated in yearly intervals for another purpose.

**The achievement age scores corresponding to the comprehension point scores were obtained by means of Table X. The unit of the point score is relatively large. Therefore, precise statements of equivalence were not shown in this table because fractional point scores are never obtained for individual pupils. This accounts for some of the apparent irregularities in the age scores for comprehension of silent reading.

In Figure 5, the median point scores derived from the arithmetic scale in Decatur are represented graphically by small crosses. The grade norms are represented by small circles. In arriving at the general relationship which exists between achievement and mental age it is necessary to bear in mind that the groups in the extremes of Table VIII are small and perhaps not representative. The pupils whose mental ages fall between 5 and 6 years probably have not learned to read well. Since the intelligence scale requires that the pupil be able to read, these mental ages are probably lower than they should be. At the upper end of the table we, of course, have very bright pupils, and bright pupils are likely to make higher scores than average or dull pupils of the same age. In Figure 5, a line has been drawn which is judged to represent the relation between achievement in the operations of arithmetic and mental age. This line can be used as a means for translating point scores into achievement ages. No simple rule can be stated as in the case of the intelligence scale. The achievement ages corresponding to the point scores yielded by the arithmetic scale are given in Table X.

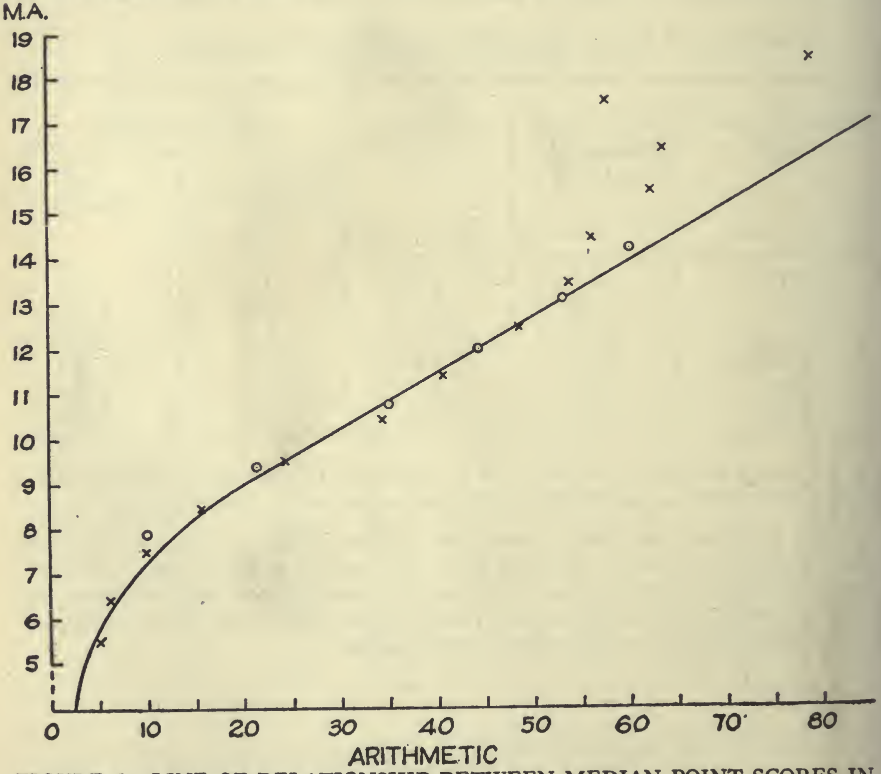


FIGURE 5. LINE OF RELATIONSHIP BETWEEN MEDIAN POINT SCORES IN ARITHMETIC AND MENTAL AGE.

TABLE X. CORRESPONDING ACHIEVEMENT AGE FOR POINT SCORES IN ARITHMETIC AND SILENT READING

Point scores			Achievement Age
Rate	Comprehension	Arithmetic	
250		140	24–0
247		136	23–6
244		132	23–0
241		128	22–6
238		124	22–0
235		120	21–6
232	20	116	21–0
229		112	20–6
226		108	20–0
222	19	104	19–6
218		100	19–0
214	18	96	18–6
210		92	18–0
206	17	88	17–6
202		84	17–0
198	16	80	16–6
194		76	16–0
190	15	72	15–6
186		68	15–0
182	14	64	14–6
178		60	14–0
174	13	56	13–6
170	12	52	13–0
165		48	12–6
160	11	44	12–0
154		40	11–6
148	10	36	11–0
141		32	10–6
133	9	28	10–0
124	8	24	9–6
113	7	20	9–0
100	6	15	8–6
85	4	10	8–0
67	3	5	7–6
47	1	0	7–0
25	0		6–6
0			6–0

Figure 6 and Figure 7 represent corresponding data for comprehension and rate of silent reading. The point scores corresponding to the various achievement ages are also given in Table X.

When measures of both achievement and general intelligence are expressed in terms of ages they may be considered comparable. The Illinois Examination provides for comparison by dividing the pupil's achievement age by his mental age.* The quotient is called his achievement quotient.(A. Q.)

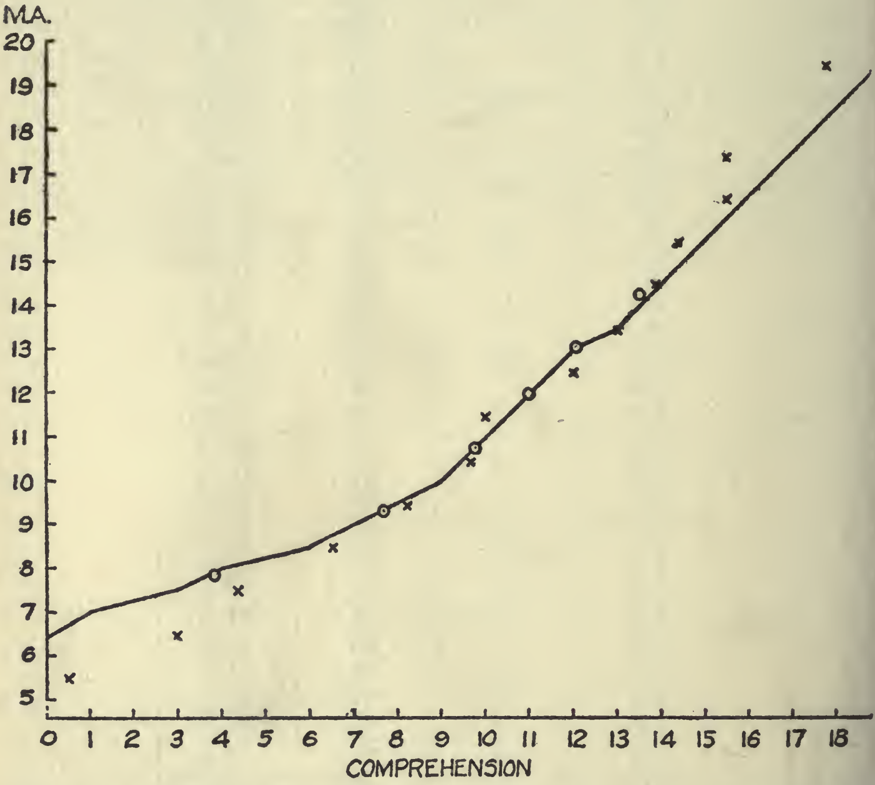


FIGURE 6. LINE OF RELATIONSHIP BETWEEN MEDIAN POINT SCORES IN COMPREHENSION OF SILENT READING AND MENTAL AGE.

---

*The achievement age norm for a pupil is numerically the same as his mental age. Hence, this procedure may also be thought of as being a comparison of a pupil's achievement with the norm for his mental age.

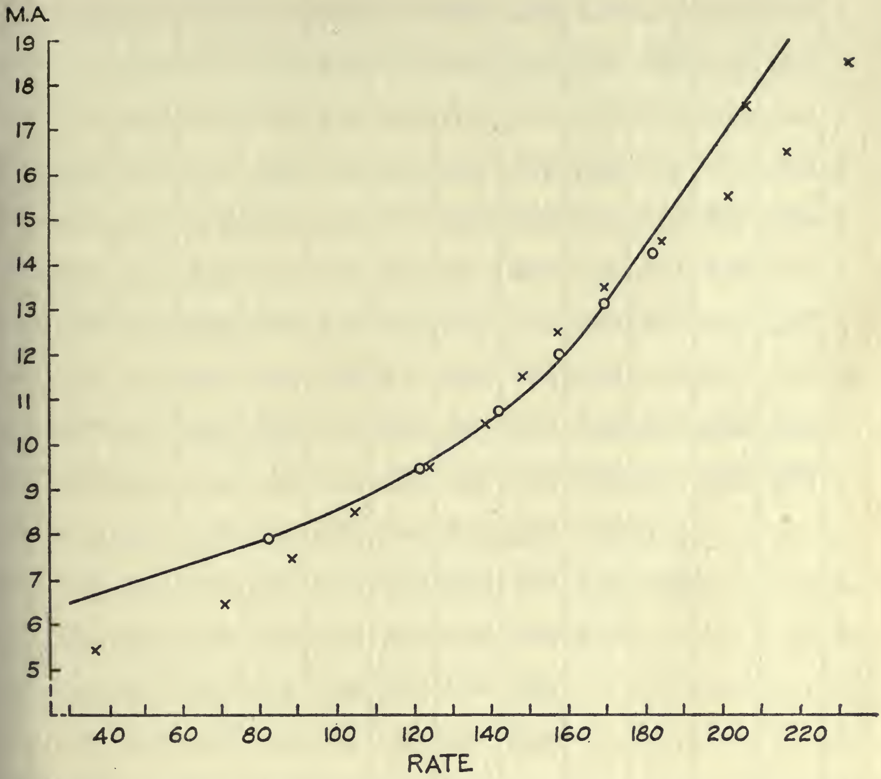


FIGURE 7. LINE OF RELATIONSHIP BETWEEN MEDIAN POINT SCORES IN RATE OF SILENT READING AND MENTAL AGE.

This is an index of the extent to which his achievement corresponds to his general intelligence or capacity to achieve. For example, if a pupil has an achievement age of 12 years and a mental age of 9 years, his achievement quotient is 133. (The decimal point is omitted as in the case of intelligence quotients.) If this pupil's mental age had been 14 years, his achievement quotient would have been 86. An achievement quotient of 100 means that the pupil has achieved exactly the average of pupils of his mental age, or that he is just up to the norm for his mental age. If his achievement quotient is 130, he has achieved thirty percent more than the average of the pupils of his mental age; on the other hand, if his achievement quotient is 75, we have evidence that he has achieved only seventy-five percent as much as the average of pupils of his mental age.

TABLE XI—ACHIEVEMENT QUOTIENTS

Rate	Point Scores		Achievement Age	Mental Age																			
	Comprehension	Arithmetic		4–6	5–0	5–6	6–0	6–6	7–0	7–6	8–0	8–6	9–0	9–6	10–0	10–6	11–0	11–6	12–0	12–6	13–0	13–6	14–0
259		152	25–6											264	251	240	229	219	210	202	194	187	181
256		148	25–0											259	246	235	224	215	206	198	191	183	177
253		144	24–6											254	241	230	220	211	202	194	187	180	174
250	23	140	24–0											249	237	226	216	206	198	190	183	176	170
247		136	23–6											244	232	221	211	202	194	186	179	173	167
244	22	132	23–0											238	227	216	207	198	190	182	175	169	163
241		128	22–6									260	246	233	222	212	202	194	186	178	172	165	160
238	21	124	22–0									254	241	228	217	207	198	189	182	175	168	162	156
235		120	21–6									249	235	223	212	202	193	185	178	171	164	158	153
232	20	116	21–0								258	243	230	218	207	198	189	181	173	167	160	155	149
229		112	20–6								251	237	224	213	202	193	184	177	169	163	157	151	146
226		108	20–0								245	231	219	208	198	188	180	172	165	159	153	147	142
222	19	104	19–6						272	255	239	226	214	203	193	184	176	168	161	155	149	144	139
218		100	19–0						265	241	233	220	208	197	188	179	171	164	157	151	145	140	135
214	18	96	18–6						258	242	227	214	203	192	183	174	167	160	153	147	142	136	132
210		92	18–0					270	252	235	221	209	197	187	178	170	162	155	149	143	138	133	128
206	17	48	17–6					262	245	229	215	203	192	182	173	165	158	151	145	139	134	129	125
202		84	17–0					255	238	223	209	197	186	177	168	160	153	147	141	135	130	125	121
198	16	80	16–6				268	248	231	216	203	191	181	172	163	156	149	143	137	131	126	122	118
194		76	16–0				260	241	224	210	197	186	176	167	159	151	144	138	133	127	123	118	114
190	15	72	15–6				252	233	217	203	191	180	170	162	154	147	140	134	129	124	119	115	111
186		68	15–0			265	244	226	210	197	185	174	165	156	149	142	136	130	124	120	115	111	107
182	14	64	14–6			256	236	219	203	190	179	169	159	151	144	137	131	126	120	116	111	107	104
178		60	14–0			247	228	211	196	184	173	163	154	146	139	133	127	121	116	112	108	104	100
174	13	56	13–6		263	239	220	204	189	177	167	157	149	141	134	128	122	117	112	108	104	100	96
170	12	52	13–0		252	230	212	196	183	171	161	151	143	136	129	123	118	113	108	104	100	96	93
165		48	12–6		243	221	204	189	176	165	155	146	138	131	124	119	113	109	104	100	96	93	89
160	11	44	12–0	258	233	213	196	181	169	158	148	140	132	126	120	114	109	104	100	96	92	89	86
154		40	11–6	247	204	204	188	174	162	152	142	134	127	121	115	109	104	100	96	92	89	85	82
148	10	36	11–0	237	214	195	180	167	155	145	136	129	122	115	110	105	100	96	92	88	85	82	79
141		32	10–6	226	205	187	172	159	148	139	130	123	116	110	105	100	96	91	88	84	81	78	75
133	9	28	10–0	216	195	178	164	152	141	132	124	117	111	105	100	95	91	87	84	80	77	75	72
124	8	24	9–6	205	186	170	156	144	134	126	118	111	105	100	95	91	87	83	80	76	74	71	68
113	7	20	9–0	195	176	161	148	137	127	119	112	106	100	95	90	86	82	79	76	73	70	67	65
100	6	15	8–6	184	167	152	140	130	120	113	106	100	95	90	85	81	78	74	71	69	66	64	61
85	4	10	8–0	174	157	143	132	122	114	106	100	94	89	85	80	77	73	70	67	65	62	60	58
67	3	5	7–6	163	148	134	124	115	107	100	94	89	84	69	76	72	69	66	63	61	58	56	54
47	1	0	7–0	153	138	126	116	107	100	94	88	83	78	74	71	67	64	62	59	57	55	53	51
25	0		6–6	142	129	117	108	100	93	87	82	77	73	69	66	63	60	57	55	53	51	49	47

Point Scores				Mental Age																			
Rate	Comprehension	Arithmetic	Achievement Age	14–6	15–0	15–6	16–0	16–6	17–0	17–6	18–0	18–6	19–0	19–6	20–0	20–6	21–0	21–6	22–0	22–6	23–0	23–6	24–0
259		152	25–6	175	169	163	158	154	149	144	141	137	134	130	127	124	121	118	116	113	111	108	106
256		148	25–0	171	166	160	155	151	146	142	139	135	131	127	125	122	119	116	114	111	109	106	104
253		144	24–6	167	162	157	152	148	143	139	136	132	129	125	122	119	116	114	111	109	106	104	102
250	23	140	24–0	164	159	154	149	145	141	137	133	129	126	123	120	117	114	112	109	107	104	102	100
247		136	23–6	161	156	151	146	142	138	134	130	127	123	120	117	114	112	109	107	104	102	100	98
244	22	132	23–0	157	152	148	143	139	135	131	127	124	121	118	115	112	109	107	105	102	100	98	96
241		128	22–6	154	149	144	140	136	132	128	125	121	118	115	112	110	107	105	102	100	98	96	94
238	21	124	22–0	151	146	141	137	133	129	125	122	119	116	113	110	107	105	102	100	98	96	94	92
235		120	21–6	147	143	138	134	130	126	123	119	116	113	110	107	105	102	100	98	96	94	92	90
232	20	116	21–0	144	139	135	131	127	123	120	116	113	110	108	105	102	100	98	96	93	91	89	88
229		112	20–6	141	136	132	128	124	120	117	114	111	108	105	102	100	98	95	93	91	89	87	86
226		108	20–0	137	133	129	125	121	117	114	111	108	105	103	100	98	95	93	91	89	87	85	84
222	19	104	19–6	134	130	125	122	118	114	111	108	105	103	100	98	95	93	91	89	87	85	83	81
218		100	19–0	130	126	122	118	115	112	108	105	103	100	97	95	93	91	89	87	85	83	81	79
214	18	96	18–6	127	123	119	115	112	109	106	103	100	97	95	93	90	88	86	84	82	81	79	77
210		92	18–0	124	120	116	112	109	106	103	100	97	95	92	90	88	86	84	82	80	78	77	75
206	17	88	17–6	120	116	113	109	106	103	100	97	95	92	90	88	86	84	82	80	78	76	75	73
202		84	17–0	117	113	110	106	103	100	97	94	92	90	87	85	83	81	79	78	76	74	73	71
198	16	80	16–6	114	110	106	103	100	97	94	92	89	87	85	83	81	79	77	75	74	72	71	69
194		76	16–0	110	107	103	100	97	94	92	89	87	84	82	80	78	76	75	73	71	70	68	67
190	15	72	15–6	107	103	100	97	94	91	89	86	84	82	80	78	76	74	72	71	69	68	66	65
186		68	15–0	103	100	97	94	91	88	86	84	81	79	77	75	73	72	70	69	67	66	64	63
182	14	64	14–6	100	97	94	91	88	86	83	81	79	77	75	73	71	69	68	66	65	63	62	61
178		60	14–0	97	93	90	88	85	83	80	78	76	74	72	70	69	67	66	64	63	61	60	59
174	13	56	13–6	93	90	87	85	82	80	77	75	73	71	70	68	66	65	63	62	60	59	58	57
170	12	52	13–0	90	87	84	82	79	77	75	73	71	69	67	65	63	62	61	60	58	57	56	55
165		48	12–6	86	84	81	78	76	74	72	70	68	66	65	63	61	60	58	57	56	55	54	52
160	11	44	12–0	83	80	78	75	73	71	69	67	65	64	62	61	59	58	56	55	54	53	52	50
154		40	11–6	80	77	75	72	70	68	66	64	63	61	59	58	57	55	54	53	52	51	49	48
148	10	36	11–0	76	74	71	69	67	65	63	62	60	58	57	56	55	53	52	51	49	48	47	46
141.		32	10–6	73	70	68	66	64	62	61	59	57	56	54	53	52	51	49	48	47	46	45	44
133	9	28	10–0	69	67	65	63	61	59	58	56	55	53	52	51	49	48	47	46	45	44	43	42
124	8	24	9–6	66	64	62	60	58	57	55	53	52	51	49	48	47	46	45	44	43	42	41	40
113	7	20	9–0	63	61	59	57	55	54	52	51	49	48	47	46	45	44	43	42	41	40	39	38
100	6	15	8–6	59	57	56	54	52	51	49	48	47	45	44	43	42	41	40	39	38	38	37	36
85	4	10	8–0	56	54	52	51	49	48	46	45	44	43	42	41	40	39	38	37	36	35	35	34
67	3	5	7–6	53	45	49	48	46	45	44	42	41	40	39	38	37	36	36	35	34			
47	1	0	7–0	49	48	46	45	43	42	41	40	39	38	36	36	35							
25	0		6–6	46	44	43	42	40	39	38	37	36											
0			6–0	42	41	40	38	37	36	35													
				14–6	15–0	15–6	16–0	16–6	17–0	17–6	18–0	18–6	19–0	19–6	20–0	20–6	21–0	21–6	22–0	22–6	23–0	23–6	24–0

Table XI is to be used in obtaining the achievement quotients. Take, for example, a pupil who has a rate score of 165, a comprehension score of 10, and an intelligence score of 60. An intelligence score of 60 is equivalent to a mental age of 11-6. (See Table VII). A rate score of 165 is equivalent to an achievement age of 12-6, and a comprehension score of 10 is equivalent to an achievement age of 11-0. To find the achievement quotient for rate of silent reading, find at the left of the table the line for an achievement age of 12-6 years. At the top of the table, find the column for a mental age of 11-6; follow down this column until you come to the line for 12-6. The number at this intersection point is the achievement quotient for rate. It is 109. To find the achievement quotient for comprehension, follow down the same column until you come to the line for an achievement age of 11-0; the number at this intersection, 96, is the achievement quotient for comprehension.

Mental age norms versus chronological age norms. In determining a basis for translating point scores into age scores we have grouped the pupils according to their mental ages as given by the Illinois General Intelligence Scale. A grouping according to chronological age would have been more convenient. Theoretically, the same results should have been obtained because the average mental age of an unselected chronological age group is identical with the average chronological age of that group. The determining motive for devising a procedure for translating point scores into age scores was to secure a convenient method for obtaining an achievement quotient. This quotient expresses the relation between a pupil's achievement and his individual achievement norm. A pupil's capacity to learn or his mental age is a much more potent factor in determining his achievement than is his chronological age. In interpreting achievement we must compare it with a pupil's capacity to learn. Hence, logically, a grouping of pupils according to mental age is to be preferred. However, since the same numerical results will be obtained from a chronological age grouping there is practical justification for using it because of the greater convenience.

Justification of disregarding school grade in determining achievement ages. In determining the median point score of pupils having the same mental age no attention was given to the school grade to which the pupils belong. Pupils in the third grade were grouped with pupils in the eighth, providing they had the same mental age. This procedure implies that in judging a pupil's achievement his school grade may be disregarded. In other words, his achievement depends only or largely upon his mental age and upon the effectiveness of the instruction received, and not on the quantity of the instruction as measured by the grade attained. The validity of this assumption was investigated by calculating separately for each school grade the median achievement age scores for each mental age group. Figure 8 represents graphically these median achievement age scores for arithmetic in the Decatur schools. It will be noticed here that there is, in general, a distinct increase from grade to grade in the median point score of pupils belonging to the same mental age. Thus, in this case our assumption is not in entire agreement with the facts. However, convenience of use demands that we have only one rule for translating point scores into achievement ages. If a different one were used for each half grade or for each grade the inconvenience and confusion arising would decidedly limit the usefulness of the scale. Furthermore, any injustice which may be done to pupils in the lower grades may be compensated for in the interpre-

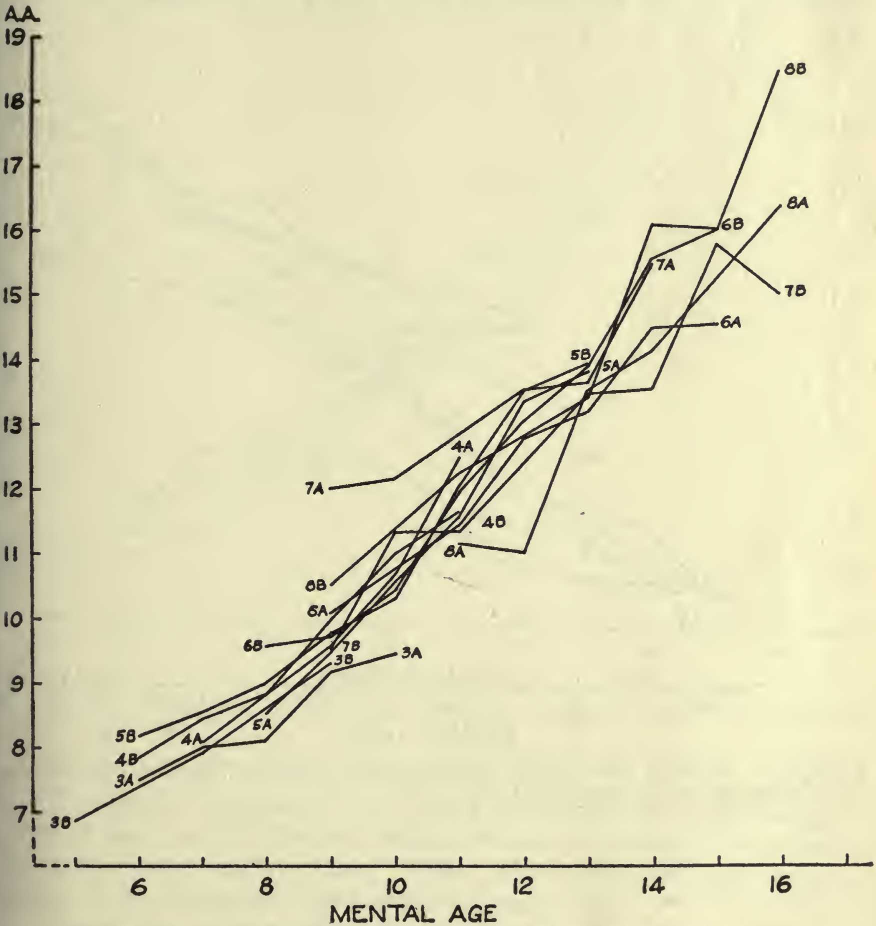


FIGURE 8. EFFECT OF SCHOOL GRADE UPON ACHIEVEMENT IN ARITHMETIC.

tation of the achievement quotients. The grade norms, however, show that in general the increase from grade to grade is slight. (See page 66).

Figures 9 and 10 report similar data for comprehension and rate of silent reading. In both of these figures it is practically impossible to distinguish between the lines representing the median achievements in the lower grades and those representing the median achievements in the upper grades. Consequently, we may say that our assumption in the case of silent reading that we disregard the placement of the pupil in translating his point score into an achievement age score is closely in agreement with the facts. Another interpretation of these figures is that ability to read silently correlates highly with general intelligence.

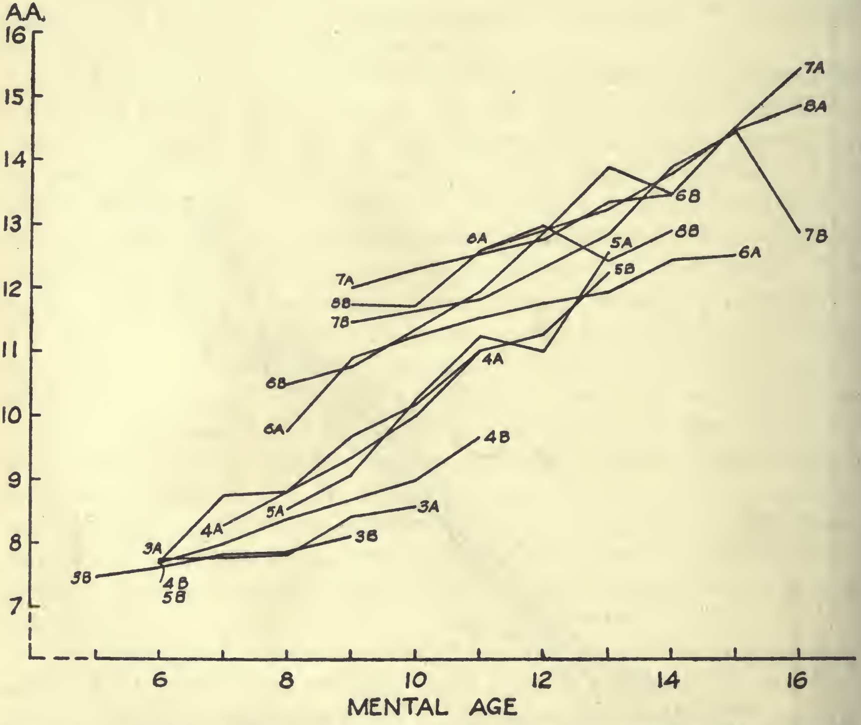


FIGURE 9. EFFECT OF SCHOOL GRADE UPON ACHIEVEMENT IN COMPREHENSION OF SILENT READING.

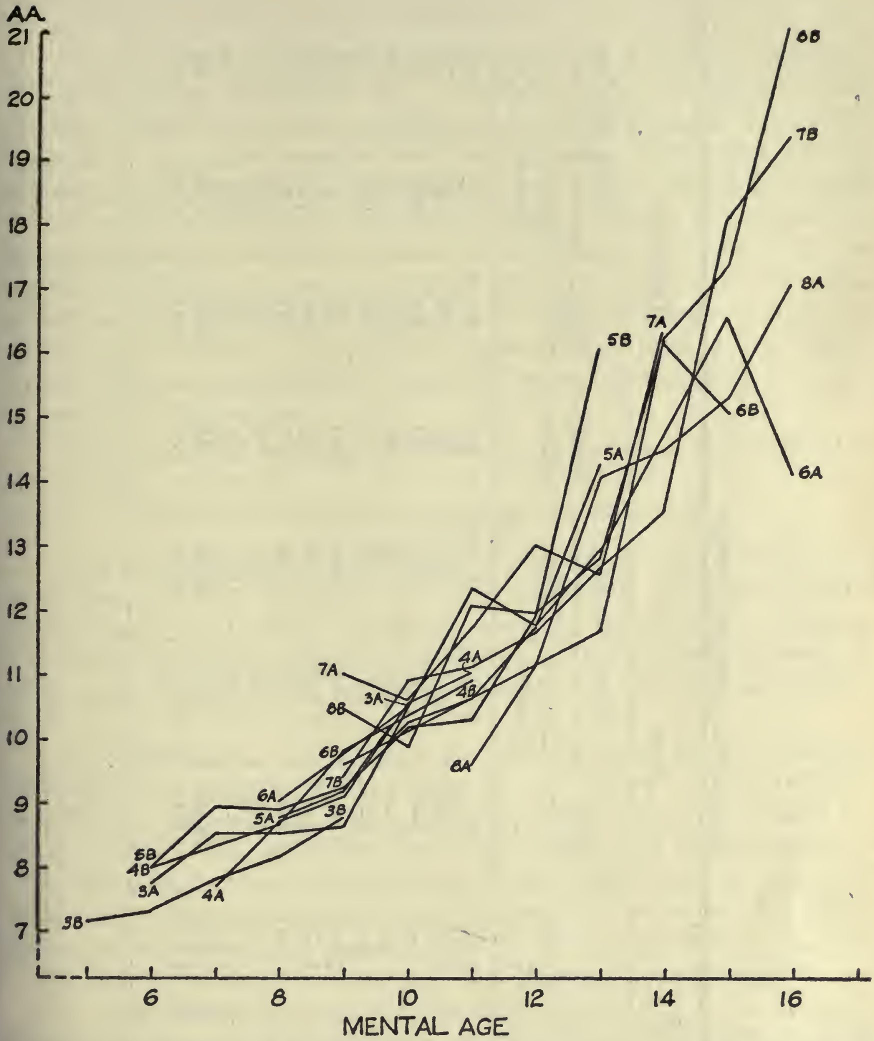


FIGURE 10. EFFECT OF SCHOOL GRADE UPON ACHIEVEMENT IN RATE OF SILENT READING.

The Illinois Examination yields four age scores and four quotient scores for each pupil. Provision has been made to calculate an average score in the case of silent reading. When this is done each pupil has ten scores. Table XII gives representative scores.

TABLE XII. TYPICAL SCORES

Pupil	Grade	Ages				Quotients			
		Mental	Arithmetic	Silent Reading		Intelligence	Arithmetic	Silent Reading	
				Comprehension	Rate			Comprehension	Rate
1	8A	12–0	10–0	11–0	11–0	83	84	92	92
2	8B	12–0	12–0	9–6	9–0	89	100	80	76
3	5A	9–0	8–0	9–0	8–0	90	89	100	89
4	3B	7–6	7–6	8–6	7–6	89	100	113	100
5	3A	6–6	7–0	6–6	6–0	73	107	100	93
6	6A	12–0	11–0	9–6	14–0	100	92	80	116
7	6B	11–6	12–0	13–0	11–6	100	104	113	100
8	4A	8–0	8–0	10–0	10–6	73	100	124	130
9	4B	7–6	7–6	8–0	7–6	79	100	106	100
10	5B	11–0	10–6	13–0	10–6	115	96	118	96
11	7A	12–6	14–0	9–6	10–6	93	112	76	84
12	7B	15–0	15–0	22–0	21–0	124	100	146	139

IV. FUNCTION.

The function of the three scales which make up the Illinois Examination is implied in their structure. The Illinois General Intelligence Scale provides a measure of general intelligence of children in grades III to VIII, inclusive.* Monroe's Standardized Silent Reading Tests, Revised, are intended to yield measures of the ability to read silently simple descriptive and narrative material when the reading is done for the purpose of answering questions. Monroe's General Survey Scale in Arithmetic is designed to yield general measures of a pupil's ability to perform the operations of arithmetic. It should be noted that the function of these tests is general rather than diagnostic. It is possible to use the sub-tests of the General Survey Scale in Arithmetic as diagnostic tests although they were not designed for this purpose.

The function of the Illinois Examination as a whole is indicated in the introductory statement.

V. VALIDITY

The ideal procedure to be followed in studying the truthfulness of the measures yielded by the scales included in the Illinois Examination would be to compare them with true measures secured by other means. However, in no case are such true measures available. It is, therefore, necessary to study the validity of these scales by methods which are obviously imperfect.

1. **Objectivity.** The scales of the Illinois Examination are highly objective with respect to the scoring of test papers. Except when a pupil fails to follow directions no questions, concerning which answers are correct, arise. The administration of the tests is also highly objective. The directions for examiners have been found to be adequate and in all cases the examiner is told very explicitly what he is to say to the pupils. Much of the explanation is also printed on the test booklet so that the pupil has an opportunity to read as well as to hear the explanation.

2. **Reliability****. In order to study the reliability of the three scales which make up the Illinois Examination the different forms were given to the same pupils. The instruction to those cooperating in this study was to give all the forms within the same half day. The scores on the different forms were compared by means of the Pearsonian coefficient of correlation and by means of other statistical devices which will be explained in the following pages.

---

*It is also recommended for use in the high school.

**A related phase of reliability is considered on page 68 under the head of "Practice effect when test is repeated."

The coefficient of correlation merely indicates the relationship between two sets of scores. It is simply an index of the extent to which the pupils make the same score on the second trial of a test that they make upon the first trial when the practise effect is disregarded. In Figure 11, we represent graphically the scores made by the fifth grade pupils on Forms 1 and 2 of the Illinois General Intelligence Scale. The coefficient of correlation for this group of scores is .92 ± .006. It is obvious that in some cases the pupils make the same score or approximately the same score on second trial. In other cases there are marked differences between the scores on the two trials.

In Figure 11 the regression line, $y = 4.92 + .80\,x$, has been drawn. Perfect correlation ($r_{12} = 1.00$) would be secured if the Form 1 scores were changed so that all points would fall upon this regression line. This would require a vertical shifting of the points. Those above would be moved downward, while those below would be moved upward. For a few of the points vertical lines have been drawn in to indicate the amount of shifting necessary. Perfect correlation would be secured if this were done with reference to any line but this regression line is the one for which the standard deviation of the shifting is the least. The other regression equation, $x = 4.69 + 1.05\,y$, possesses similar properties for a horizontal shifting.

The amounts of changes necessary to secure perfect correlation may be thought of as departures from perfect correlation. The magnitude of these changes is described by the equation for the probable error of estimate,

$$\text{P. E.}_{\text{est}} = .6745\,\sigma y \sqrt{1 - r^2_{12}}.$$

Substituting in this equation for σy and r_{12} we have the probable error of estimate equal to 6.06. The probable error of estimate is more easily interpreted as the index of the degree of correlation that exists than as the coefficient of correlation.

Since neither set of scores gives accurate measures of intelligence the differences between the pairs of scores do not truthfully represent the degree of inaccuracy of either set of scores. The error of any score is the difference between it and a pupil's true score. We may define a true score as the average of an infinite number of scores after they have been corrected for practise effect, fatigue, and other factors which would tend to increase or decrease the averages of the successive sets of scores. Such true scores are obviously not obtainable. It is, however, possible to determine the coefficient of correlation between either set of obtained scores and the corresponding true scores. This is done by the formula,*

$$r_{1t}=\sqrt{r_{12}}.$$

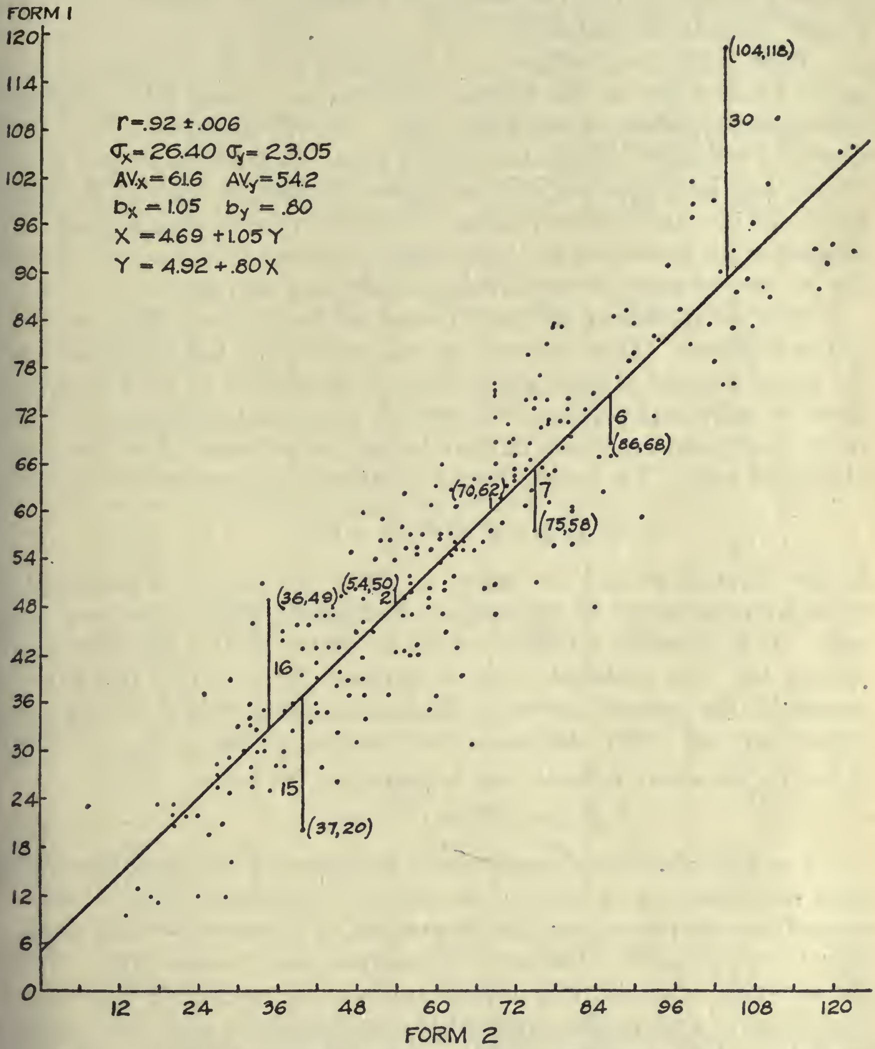


FIGURE 11. CORRELATION OF FORM 1 SCORES WITH FORM 2 SCORES OF THE ILLINOIS GENERAL INTELLIGENCE SCALE, FIFTH GRADE.

In this formula r_{12} is the coefficient of correlation between the two sets of obtained scores and r_{1t} is the coefficient of correlation between one set of obtained scores and the corresponding true scores. To distinguish the

---

*See Kelley, T. L. "A simplified method of using scaled data for purposes of testing," *School and Society*, 4:74, July 8, 1916, and "The reliability of test-scores," *Journal of Educational Research*, 3:370-79, May, 1921.

coefficient of correlation of a set of obtained scores with the corresponding set of true scores from the coefficient of correlation between two sets of obtained scores we call the latter the *coefficient* of reliability and the former the *index* of reliability.

Table XIII gives coefficients of reliability (r_{12}) for each of the three scales which make up the Illinois Examination. Table XIV gives the corresponding indices of reliability (r_{1t}). It will be noted that a high degree of reliability is indicated in most cases. In some instances it is unusually high in comparison with the degree of reliability reported for other tests. In the case of silent reading certain instructions were not followed by some of the examiners, and it is thought that their failure to do so caused the two sets of scores to correlate less highly than they should.

Both the coefficient of reliability and the index of reliability are difficult to interpret. They express a general relationship but not in terms of the actual amount of error which must be allowed for in the case of the scores of individual pupils. It is possible to calculate another and more easily interpreted expression of the reliability or accuracy of the measures yielded by tests. The probable error of estimate is given by the formula,

$$\text{P. E.}_{\text{est}} = .6745\,\sigma\sqrt{1 - r^2_{1t}}.$$

In the formula σ may be taken as either σ_1 or σ_2. Theoretically, these are expected to be equivalent. Practically, slight differences may exist. It is, therefore, advisable to use as the value of σ the average of σ_1 and σ_2. The probable error of estimate (P. E. $_{\text{est}}$) in this case is essentially the probable error in the measurements yielded by the test. Hence, we call it by this name and use the symbol, P. E.$_{\text{m}}$. Since $r_{1t} = \sqrt{r_{12}}$, the above formula may be written in the form,

$$\text{P. E.}_{\text{m}} = .6745\,\sigma\sqrt{1 - r_{12}}.$$

The probable error of measurement calculated by the above formula is to be interpreted as an index of the amount of departure of the obtained scores from the true scores. In other words, it is the error which the obtained score involves. This error is described as a probable error. Such a description, of course, tells us nothing about the magnitude of the error in the case of a particular pupil but it does describe in a general way the magnitude of the errors involved in a group of scores.

TABLE XIII. RELIABILITY COEFFICIENTS

Grade	Form	General Intelligence		Arithmetic		Silent Reading		
		No. of Pupils	r_{12}	No. of Pupils	r_{12}	No. of Pupils	Rate r_{12}	Comprehension r_{12}
III	1 with 2	76	.86	299	.80	116	.69	.63
	2 with 3						.72	.64
	1 with 3						.72	.63
IV	1 with 2	120	.93	271	.86	112	.79	.63
	2 with 3						.80	.64
	1 with 3						.70	.60
V	1 with 2	243	.92	256	.88	120	.79	.69
	2 with 3						.86	.74
	1 with 3						.72	.71
III to V	1 with 2			826	.95	348	.78	.80
	2 with 3						.85	.64
	1 with 3						.76	.82
VI	1 with 2	198	.82	271	.76	139	.87	.52
	2 with 3						.89	.86
	1 with 3						.90	.70
VII	1 with 2	157	.80	257	.71	100	.72	.68
	2 with 3						.74	.75
	1 with 3						.61	.63
VIII	1 with 2	164	.67	171	.79	119	.91	.85
	2 with 3						.88	.71
	1 with 3						.91	.85
VI to VIII	1 with 2			699	.76	358	.79	.72
	2 with 3						.84	.76
	1 with 3						.75	.76
III to VIII	1 with 2	958	.92					

TABLE XIV. INDEX OF RELIABILITY

Grade	Form	General Intelligence		Arithmetic		Silent Reading		
		No. of Pupils	r_{1t}	No. of Pupils	r_{1t}	No. of Pupils	Rate r_{1t}	Comprehension r_{1t}
III	1 with 2	76	.93	299	.89	116	.83	.79
	2 with 3						.85	.80
	1 with 3						.85	.79
IV	1 with 2	120	.96	271	.93	112	.89	.79
	2 with 3						.89	.80
	1 with 3						.84	.77
V	1 with 2	243	.96	256	.94	120	.89	.83
	2 with 3						.93	.86
	1 with 3						.85	.84
III to V	1 with 2			826	.97	348	.88	.89
	2 with 3						.92	.80
	1 with 3						.87	.91
VI	1 with 2	198	.91	271	.87	139	.93	.72
	2 with 3						.94	.93
	1 with 3						.95	.84
VII	1 with 2	157	.89	257	.84	100	.85	.82
	2 with 3						.86	.87
	1 with 3						.78	.79
VIII	1 with 2	164	.82	171	.89	119	.95	.92
	2 with 3						.93	.84
	1 with 3						.95	.92
VI to VIII	1 with 2			699	.87	358	.89	.85
	2 with 3						.92	.87
	1 with 3						.87	.87
III to VIII	1 with 2	958	.96					

TABLE XV. PROBABLE ERRORS OF MEASUREMENT AND RATIO OF PROBABLE ERRORS OF MEASUREMENT TO AVERAGE SCORES

Grade	Forms	Intelligence		Arithmetic		Silent Reading			
						Comprehension		Rate	
		P.E.$_m$	$\frac{\text{P.E.}_m}{\text{Av.}}$	P.E.$_m$	$\frac{\text{P.E.}_m}{\text{Av.}}$	P.E.$_m$	$\frac{\text{P.E.}_m}{\text{Av.}}$	P.E.$_m$	$\frac{\text{P.E.}_m}{\text{Av.}}$
III	1 with 2	3.5	0.10	2.6	0.17	1.2	0.16	13.7	0.12
	1 with 3					1.3	0.17	13.7	0.11
	2 with 3					1.3	0.17	14.5	0.12
IV	1 with 2	5.5	0.09	4.6	0.12	1.4	0.16	10.3	0.08
	1 with 3					1.2	0.14	12.8	0.10
	2 with 3					1.2	0.13	10.9	0.09
V	1 with 2	4.7	0.08	4.4	0.09	1.2	0.10	12.0	0.08
	1 with 3					0.9	0.07	13.7	0.08
	2 with 3					1.1	0.09	10.7	0.07
III to V	1 with 2			3.2	0.10	1.0	0.11	13.1	0.10
	1 with 3					1.0	0.10	13.7	0.10
	2 with 3					0.9	0.09	11.5	0.08
VI	1 with 2	5.5	0.07	6.3	0.10	1.3	0.10	9.1	0.05
	1 with 3					1.1	0.09	8.2	0.05
	2 with 3					1.1	1.08	9.5	0.05
VII	1 with 2	6.4	0.07	5.3	0.09	1.2	0.09	13.6	0.07
	1 with 3					1.4	0.10	17.0	0.09
	2 with 3					1.1	0.07	14.9	0.07
VIII	1 with 2	7.7	0.08	5.4	0.08	0.8	0.05	7.5	0.04
	1 with 3					0.9	0.06	7.5	0.04
	2 with 3					1.1	0.07	9.8	0.05
VI to VIII	1 with 2			6.2	0.10	1.1	0.08	12.0	0.07
	1 with 3					1.1	0.08	13.7	0.07
	2 with 3					1.1	0.08	11.4	0.06
III to VIII	1 with 2	5.3	0.07						

The probable error of measurement tends to increase as the scores become larger and the significance of an error depends upon the magnitude of the score with which it is associated. For this reason added meaning can be given to our description of the errors by calculating the ratio of the probable error of measurement to the average score. This gives the probable error of measurement in the form of a percent of the score. In Table XV, the probable errors of measurement and the ratios of these to the average scores are given. The numbers of pupils involved are the same as those given in Tables XIII and XIV. A probable error of measurement of 3.5 for the Illinois General Intelligence Scale in the third grade means that the point scores of fifty percent of the pupils will involve errors less than this amount. The remaining fifty percent of the scores will involve errors greater than 3.5. A somewhat more general statement is that, on the average, the scores obtained from third grade pupils will involve probable errors of ten percent of their magnitude. The larger the score the larger the probable error.

The magnitude of these probable errors of measurement may appear to be somewhat disturbing but they are not large in comparison with those calculated for other tests. In the case of the Illinois General Intelligence Scale the average probable error of measurement amounts to about six months. This is approximately the same as that calculated for the Stanford Revision of the Binet Scale for the Measurement of Intelligence.* An unpublished study of the reliability of a group of silent reading tests made by the writer gave ratios of reliability which are much larger in many cases.

3. **Discrimination.** The shape of the distribution of the scores which a test yields throws some light upon its validity. In order that a test be valid the scores must show differences of the traits measured when these differences exist. When a representative group of pupils is measured with reference to a mental or a physical trait we may expect to find a distribution closely approximating the normal shape. When the number of cases is large this approximation should be close if there is proper discrimination. Any marked departure from the normal shape indicates that for some pupils at least there is a lack of discrimination. On the other hand, when we have a normal distribution we cannot know definitely that our measures are accurate. The shape of the distribution, therefore, has only a negative significance. We can only say that when there is a striking departure from normality there is a lack of discrimination and, hence, inaccuracy of measurement for some pupils.

---

*Otis, A. S. and Knollin, H. E. "Reliability of Binet Scale and Pedagogical Scales." *Journal of Educational Research*, 4:121-43, September, 1921.

Figures 12, 13, 14, and 15 represent graphically the distributions of scores yielded by the scales of the Illinois Examination for certain groups of pupils. In Figure 12, the distributions with respect to mental age are given for grades IV, VI, and VIII. Each of these distributions includes several thousand children. They have been reduced to the same basis by expressing the frequencies in terms of percents. Each of these curves closely approximates the normal distribution. Thus, in the case of the Illinois General Intelligence Scale, the shape of the distributions furnishes no evidence that there is a lack of discrimination.

In Figures 13, 14, and 15, we give the distributions of the achievement ages for grades III, V, VI, and VIII, for the pupils of one city. Because the original bases for translating point scores into achievement ages were found to be incorrect it is not possible to use the scores of the pupils on which the distributions for mental age given in Figure 12 are based. The distributions of achievement ages in arithmetic (Figure 13) exhibit striking departures from the normal shape in grades III and V. In grades VI and

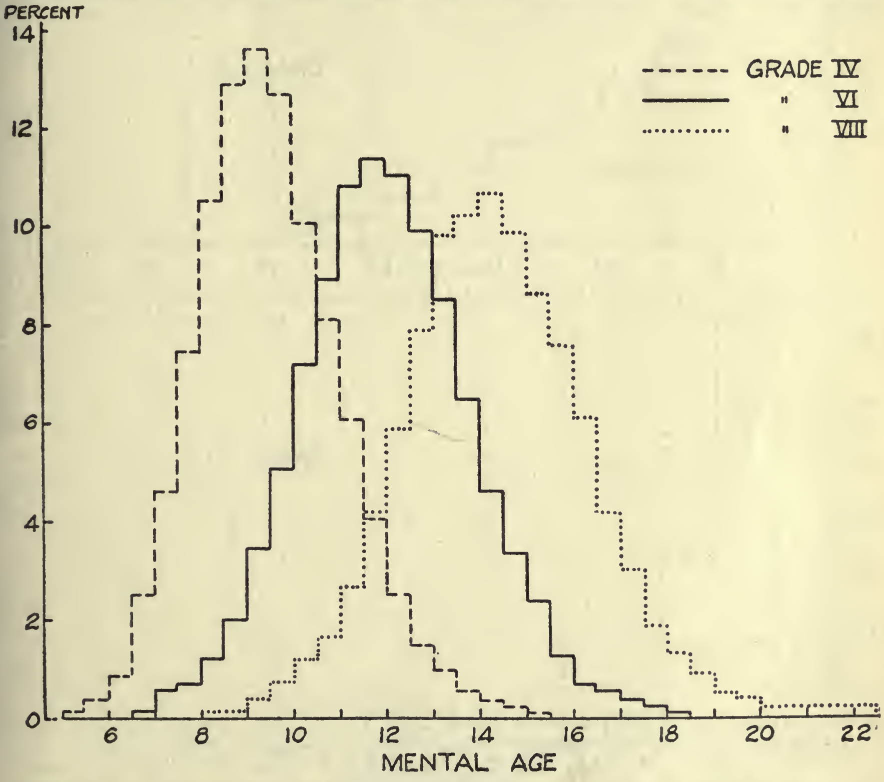


FIGURE 12. DISTRIBUTION OF PUPILS ACCORDING TO MENTAL AGE.

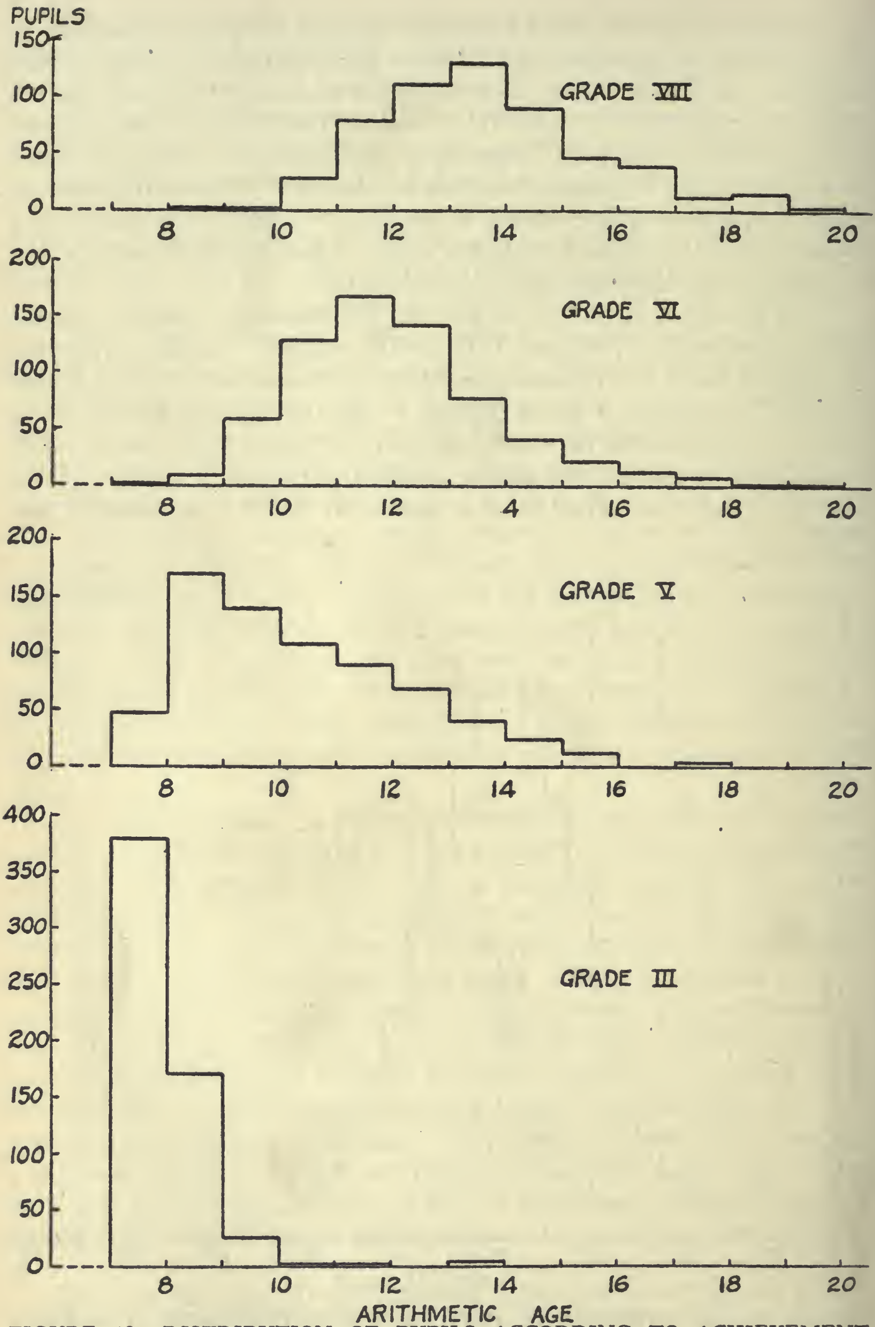


FIGURE 13. DISTRIBUTION OF PUPILS ACCORDING TO ACHIEVEMENT AGE IN ARITHMETIC.

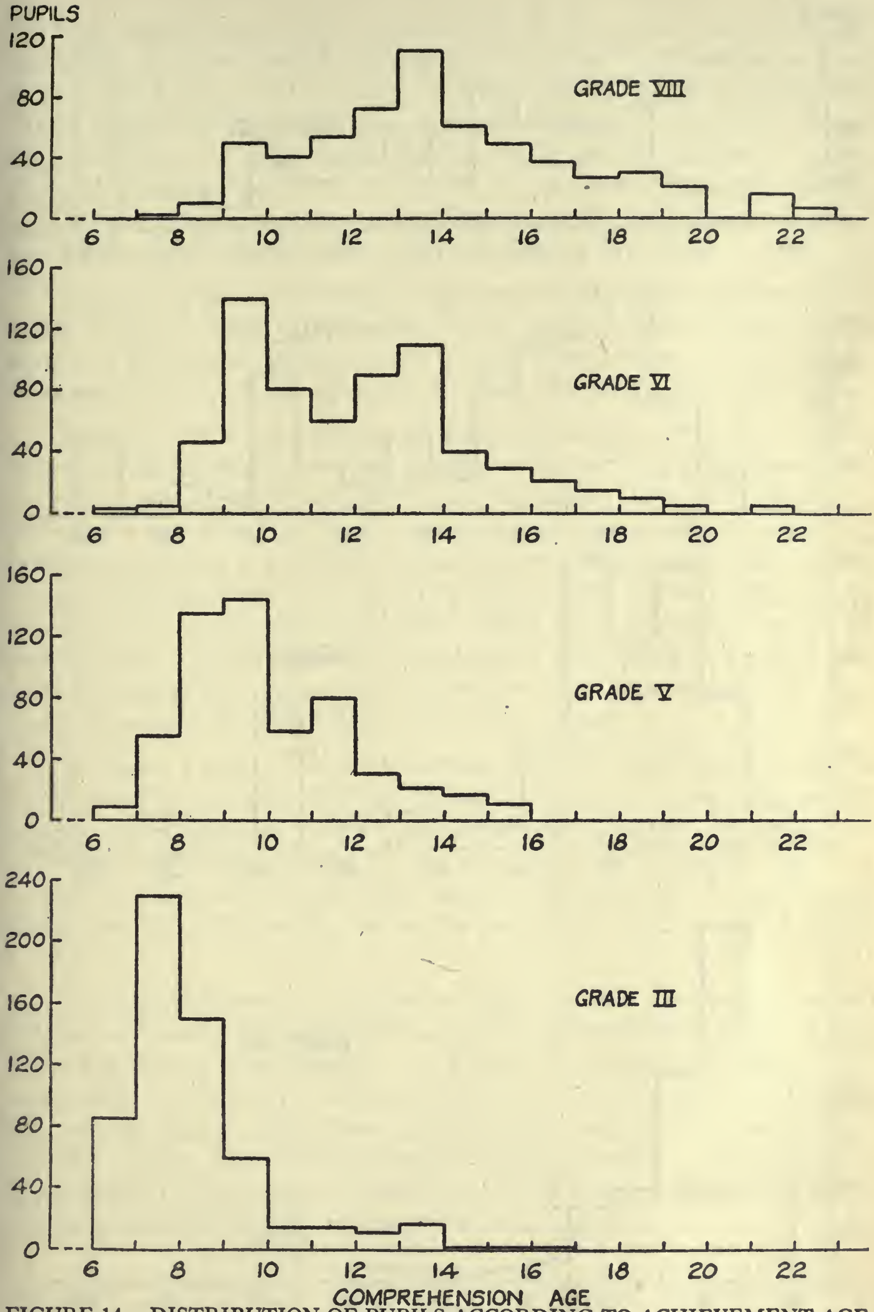


FIGURE 14. DISTRIBUTION OF PUPILS ACCORDING TO ACHIEVEMENT AGE IN COMPREHENSION OF SILENT READING.

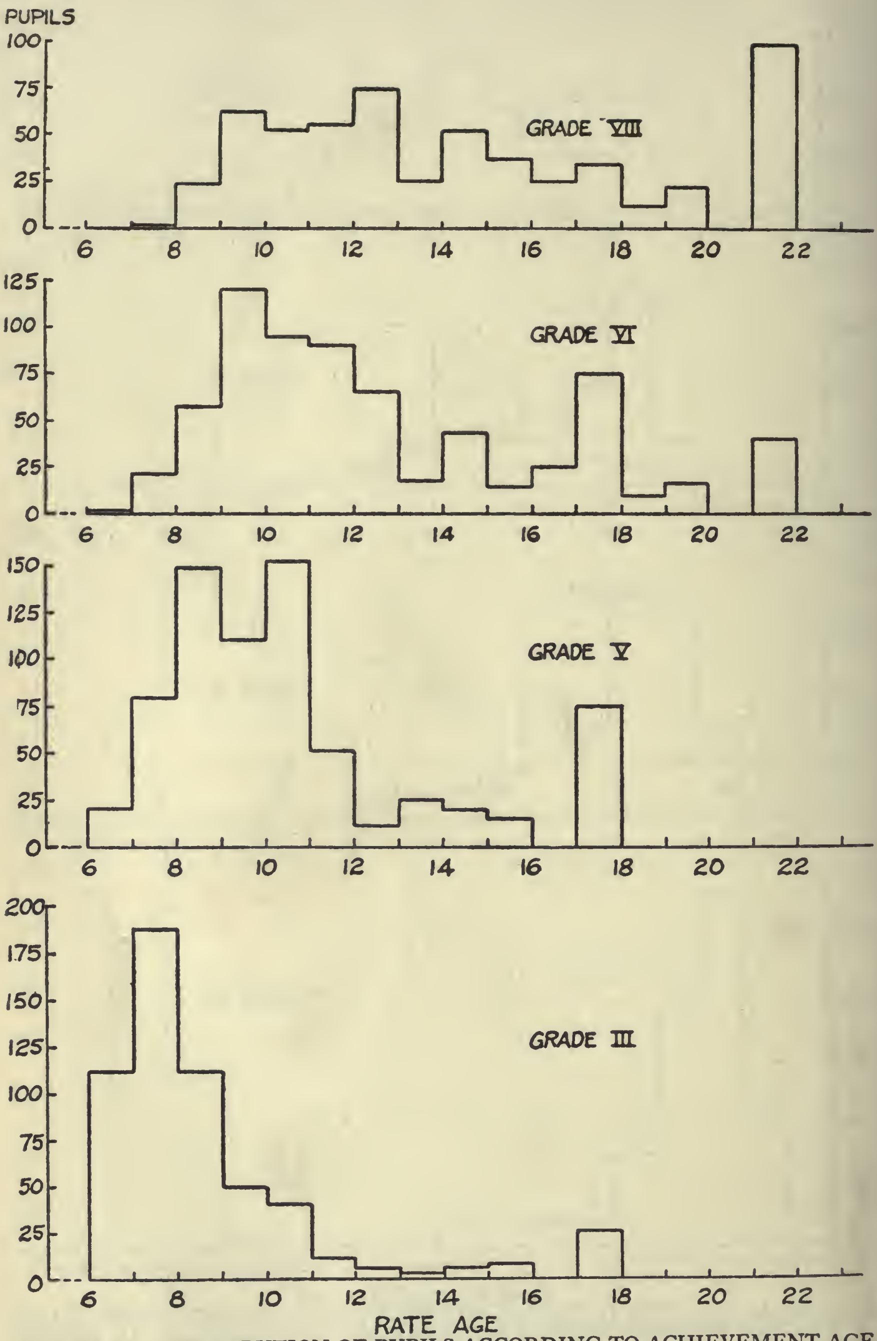


FIGURE 15. DISTRIBUTION OF PUPILS ACCORDING TO ACHIEVEMENT AGE IN RATE OF SILENT READING.

VIII the resemblance to the normal curve is very close. It will be remembered that in measuring achievement in the field of arithmetic one scale is used for grades III, IV, and V, and another for grades VI, VII, and VIII. It appears, therefore, that the scale designed for use in the lower grades does not measure accurately the arithmetical abilities of some pupils in the lower grades. The fact that such a large percent of pupils in these grades are grouped together shows that this scale fails to discriminate between some pupils having different degrees of ability.

Figure 14 gives the corresponding distributions for comprehension of silent reading. These distributions more nearly approach the normal shape. The one for the sixth grade exhibits two modes. In each grade there are a few very high scores which destroy the symmetry of the figures.

In Figure 15, we represent graphically the corresponding distributions for rate of silent reading. These graphs exhibit striking departures from the normal shape, particularly in the upper grades. It is clear, therefore, that in the case of rate of silent reading the measures must be considered lacking in accuracy in a number of cases. In the eighth grade, particularly, there is a failure to discriminate with respect to the rate of reading in the cases of many pupils who read rapidly. This tendency is also seen in the other grades. The irregularities are probably due in part to the fact that rate of reading the unconnected exercises varies from exercise to exercise. Some may be read rapidly, others very slowly.

In Figure 4 we gave the total distribution of the intelligence quotients for the pupils in grades IIIB to VIIIA, inclusive, in a large city school system. (See page 29). This distribution approximates the normal distribution very closely except for the interval from 90 to 100. The frequency in this interval is less than that for the intervals on either side. This, however, is probably due to the use of our table to calculate the I. Q.'s.

For the achievement quotients we give only the distributions for all grades combined. The distributions for the separate grade groups do not show any marked irregularities. In Figure 16, the distribution of the achievement quotients for arithmetic is given. This closely approximates the normal shape. This close approximation will also be observed in Figure 17 in the case of comprehension of silent reading. For the rate of silent reading (See Figure 18) the distribution is less symmetrical, as we might expect after noting the irregularities in Figure 15.

An examination of the distributions of the intelligence quotients and of the achievement quotients does not reveal any indication of a marked lack of discrimination except possibly in the case of the rate of silent reading. One might be inclined to conclude that there was evidence of a lack of discrimination in the case of the intelligence quotient if the distributions for mental age did not approach the normal curve as closely as they do.

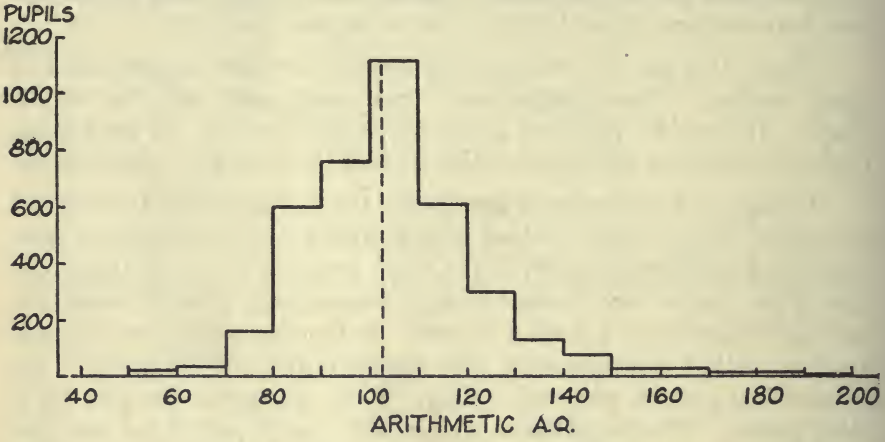


FIGURE 16. DISTRIBUTION OF PUPILS ACCORDING TO ACHIEVEMENT QUOTIENTS IN ARITHMETIC.

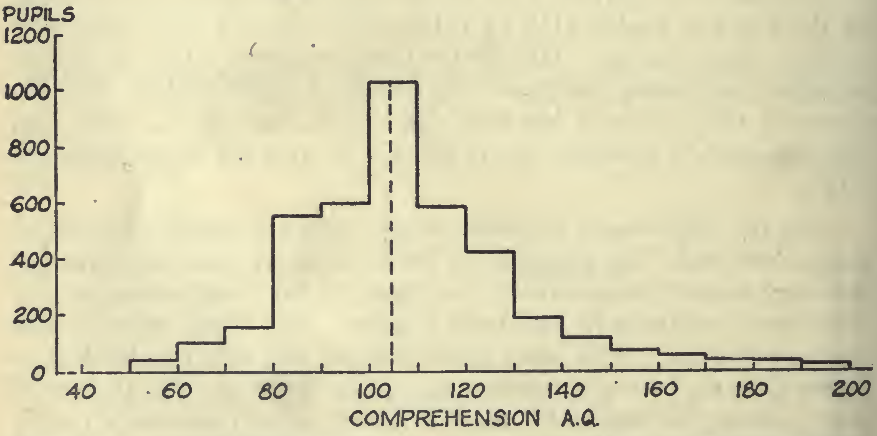


FIGURE 17. DISTRIBUTION OF PUPILS ACCORDING TO ACHIEVEMENT QUOTIENTS IN COMPREHENSION OF SILENT READING.

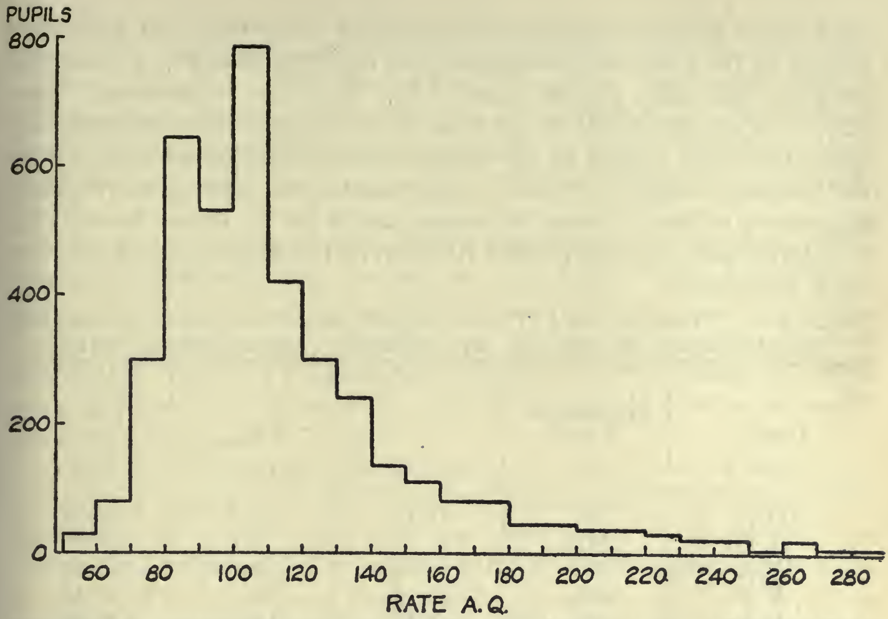


FIGURE 18. DISTRIBUTION OF PUPILS ACCORDING TO ACHIEVEMENT QUOTIENTS IN RATE OF SILENT READING.

4. **Comparison with criterion measures.** The Illinois General Intelligence Scale was given to 203 pupils whose mental ages had also been determined by the Stanford Revision of the Binet Scale for Measuring Intelligence. The correlation between the mental ages determined by these two scales is .74 ± .02. The probable error of estimate is 1.2 years. This means that in 50 percent of the cases, the mental age, as determined by the Illinois General Intelligence Scale, differed from that as determined by the Binet Scale by 1.2 years or less. This lack of agreement between the measures secured by these two scales is not due solely to errors in the measures yielded by the Illinois General Intelligence Scale. The Stanford Revision of the Binet Scale for Measuring Intelligence also yields measures which involve errors of approximately the same magnitude as those of the Illinois General Intelligence Scale. (See pages 24-27).

In November, 1920, both the Illinois General Intelligence Scale, Form 1, and the National Intelligence Scale, Form 1, were given to 3615 pupils in eight elementary schools in Chicago. The correlation between the scores obtained from these two tests is indicated in Table XVI. The probable errors of estimate indicate that the agreement is not close. The probable error of estimate, when all grades are taken together, is 11.5.

This means that the departure from perfect correlation with the scores yielded by the National Intelligence Scale is greater than 11.5 points in 50 percent of the cases, and less than 11.5 in 50 percent of the cases. Since ten points are equivalent to one year of mental age, the relationship between the scores yielded by the Illinois General Intelligence Scale and by the National Intelligence Scale is approximately the same as the relationship shown to exist between the scores yielded by the Illinois General Intelligence Scale and the Stanford Revision of the Binet Scale for Measuring Intelligence.

TABLE XVI. CORRELATION BETWEEN SCORES YIELDED BY ILLINOIS GENERAL INTELLIGENCE SCALE AND BY NATIONAL INTELLIGENCE SCALE

Grade	Number of Cases	r	$P.E._{est}$	$P.E._{est}$
				Aver.
III A	357	0.53	9.1	0.22
IV B	416	0.70	9.6	0.18
IV A	335	0.74	8.0	0.14
V B	460	0.55	8.7	0.14
V A	285	0.47	12.0	0.19
VI B	383	0.44	12.6	0.17
VI A	259	0.67	10.8	0.13
VII B	350	0.70	11.0	0.12
VII A	210	0.68	10.3	0.11
VIII B	271	0.72	10.2	0.10
VIII A	289	0.69	10.9	0.10
All Grades	3615	0.81	11.5	0.16

The Illinois General Intelligence Scale, Form 1, was given to a number of sixth-grade pupils whose I. Q.'s, as determined for the Otis Group Intelligence Tests, were available.* The coefficient of correlation for 83 VIA pupils was .82±.02. For 124 VIB pupils the value of r was .83±.02. The probable error of estimate was 6.4 in the first case and 5.9 in the latter.

The Pintner Non-Language Group Intelligence Tests are represented to have a reliability coefficient of .72. This is by mental indices and not by point scores. The two sets of measures were obtained by use of the same test after an interval of two years. The number of children tested was 46. These group intelligence tests were also given during the same semester to 300 children whose mental ages had been determined by the Stanford Revision of the Binet Scale for Measuring Intelligence. The coefficient of correlation between the point scores yielded by the Pintner Group Test and the mental ages determined by the Binet Test was .80.*

---

*The writer is indebted to Superintendent L. W. Keeler, Michigan City, Indiana for these data.

No data are available at this time for making comparison between the measures of achievement yielded by the achievement scales included in the Illinois Examination and by other similar scales. Neither are data available for comparison of measures of achievement with teachers' estimates.

5. **Inferences concerning validity based upon the structure of the test and its administration.** In the case of the Illinois General Intelligence Scale the sub-tests have frequently been used by other makers of instruments for measuring general intelligence. At the time the Illinois General Intelligence Scale was constructed a number of other intelligence scales were analyzed with reference to sub-tests and the ones most frequently found were incorporated in this scale.

The intercorrelations between the seven sub-tests were studied by choosing at random two sets of 120 test papers each. In securing these samples ten papers were chosen from each half grade. The correlation of each sub-test with each other sub-test and with the total test score is given in Table XVII. The upper number in each case refers to the first sample and the lower to the second sample. It will be noted that the correlation of a sub-test with the total score is higher than the correlation with the other sub-tests. The correlations with the total score are high in every case. Sub-tests 2 and 3 have the highest coefficients of correlation. The intercorrelations between the sub-tests are relatively high but it must be remembered that we have used a wide range of talent extending from IIIB to VIIIA. The table, however, is evidence that the overlapping between the sub-tests is not great and is approximately uniform.

It is interesting to note the differences between the corresponding coefficients yielded by the two samples. The differences are never large and are within the range indicated by the probable errors of the coefficients of correlation due to sampling.

The Illinois General Intelligence Scale is explicitly a verbal test. Ability to read is a prerequisite. For this reason it may be urged that it does not permit non-verbal elements of intelligence to function. This objection, of course, applies to other verbal tests. No data are at hand to show the limitation which this feature places upon the scale.

The silent reading test included is a revision of Monroe's Standardized Silent Reading Test. In this revised form certain features of the original test which were found unsatisfactory have been eliminated. The scoring has been made objective. The exercises are more uniform and a more

---

*Pintner, Rudolph and Marshall, Helen. "Combined mental-educational survey." *Journal of Educational Psychology*. 12:32-43, January, 1921.

TABLE XVII. INTERCORRELATIONS BETWEEN THE SUB-TESTS OF THE ILLINOIS GENERAL INTELLIGENCE SCALE, FORM I*

Sub-Tests	Sub-Tests							
	1	2	3	4	5	6	7	Total
1		.58 .68	.63 .73	.51 .48	.51 .44	.57 .60	.56 .62	.77 .84
2	.58 .68		.74 .75	.65 .60	.49 .61	.67 .73	.64 .70	.83 .89
3	.63 .73	.74 .75		.63 .61	.57 .70	.64 .72	.68 .76	.88 .93
4	.51 .48	.65 .60	.63 .61		.41 .50	.58 .55	.52 .54	.77 .78
5	.51 .44	.49 .61	.57 .70	.41 .50		.51 .51	.47 .57	.70 .77
6	.57 .60	.67 .73	.64 .72	.58 .55	.51 .51		.55 .61	.79 .80
7	.56 .62	.64 .70	.68 .76	.52 .54	.47 .57	.55 .61		.79 .86
Total Score	.77 .84	.83 .89	.88 .93	.77 .78	.70 .77	.79 .80	.79 .86	

*Two coefficients are given for each pair. Each was calculated from a random selection of ten test papers from each half grade group III B to VIII A, inclusive.

precise measurement of the rate of silent reading is secured. Experience has shown that the test is too short for the time limit allowed. This, in the case of the most fluent readers, prevents one from securing valid measures of reading ability. To one acquainted with the nature of reading ability and its measurement, this one test obviously measures in a general way, only one type of silent reading ability. To measure completely all phases of silent reading ability would require a battery of tests.

In the case of Monroe's General Survey Scale in Arithmetic the subtests are judged to represent the most important types of examples learned by pupils in the sequence of grades for which they are intended. The choice represents the judgment of the author but is not inconsistent with other groups of tests used to measure ability of pupils in the operations of arithmetic. A single general score has been used to describe the pupil's ability. This is, obviously, a composite not only of the scores of the different sub-tests but also of the dimensions of rate and accuracy. However, since a single general measure is desired this does not constitute a serious criticism of the scale.

Summary for validity. By way of summary we may say that the scales which make up the Illinois Examination compare favorably in respect to validity with our best tests. It is, however, clear that the scales possess certain limitations which should be kept in mind when the scores are interpreted.

VI. VALIDITY OF SIGNIFICANCE

Probably all readers will admit that the abilities measured by this battery of tests are important in the education of children. They are probably the abilities most often measured. The Illinois Examination includes a plan, however, for combining measures of intelligence and measures of achievement. It is, therefore, appropriate that we point out the significance of the comparison proposed.

Significance of achievement quotients. In Figure 19, the achievement ages for both arithmetic and reading average are shown for City A and City B. It is clearly evident, from this figure, that when considered

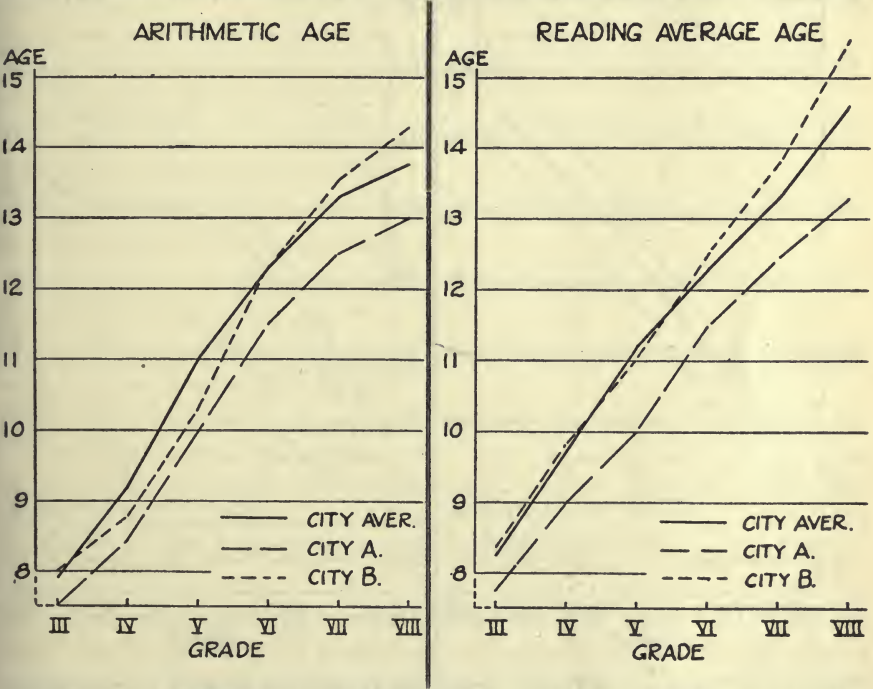


FIGURE 19. ACHIEVEMENT AGES FOR ARITHMETIC AND FOR READING (AVERAGE) FOR CITY A AND CITY B.

grade for grade, City B is distinctly superior to City A in both arithmetic and reading. The superiority is more pronounced in the case of reading. If this were the only information at hand for these two cities the conclusion would be that the schools of City B were superior to those of City A, and that City A was below the city average.

In Figure 20, the median mental age and the median I. Q. for the several grades of these two cities are represented. The school system in City B is shown to have been so organized that in each grade the median mental age of the pupils is about one year greater than in City A. The median intelligence quotients are also distinctly higher. In fact, City B, except in the third grade, is distinctly above the city average. These facts mean that, grade for grade, City B has superior pupil material. This is probably due to a difference in the organization of the school systems of the two cities. In City B, the school system is so organized that in the upper grades the pupils are highly selected. In City A, the opposite condition prevails. Therefore, we should expect City B to secure distinctly superior achievements, grade for grade.

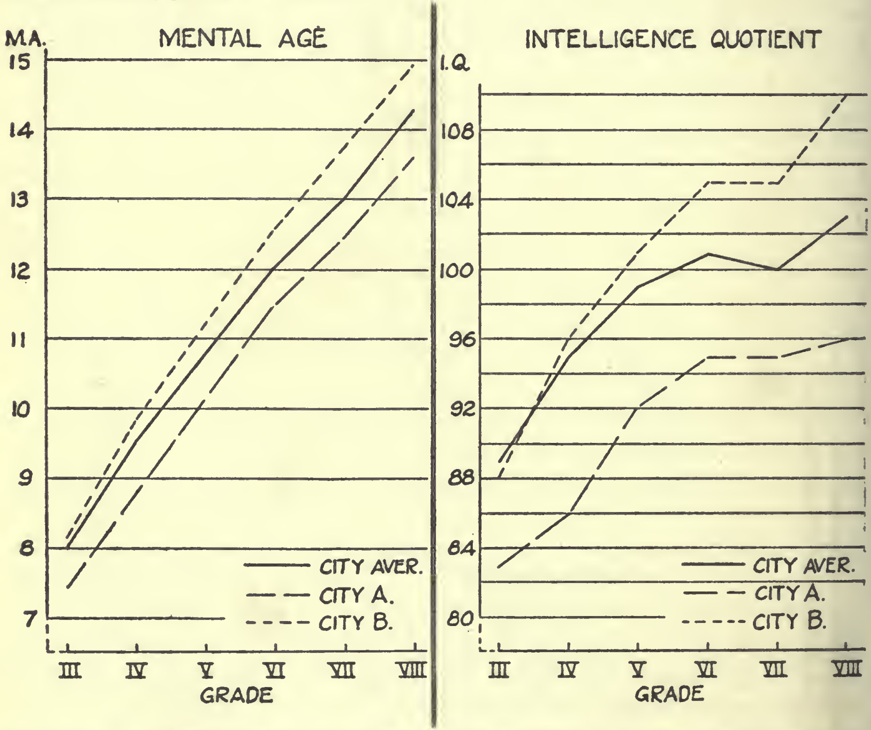


FIGURE 20. MEDIAN MENTAL AGE AND MEDIAN I. Q. FOR EACH GRADE FOR CITY A AND CITY B.

We now have to answer the question, "Is the school system of City B as effective as that of City A when the character of the pupil material in the two cities is taken into account." Figure 21 answers this question. It shows the achievement quotients for these two cities. In the case of arithmetic the quotients for City B are, except in the eighth grade, below those for City A. They are also below the city average. In the case of reading City B surpasses City A only in the seventh and eighth grades. Therefore, we must conclude that, with the exception of the eighth grade and in part the seventh, the school system of City B is less effective in the teaching of the operations of arithmetic and silent reading than the school system of City A. By using the mental ages of the pupils we have thus been able to avoid an erroneous interpretation of the measures of achievement for these two cities.

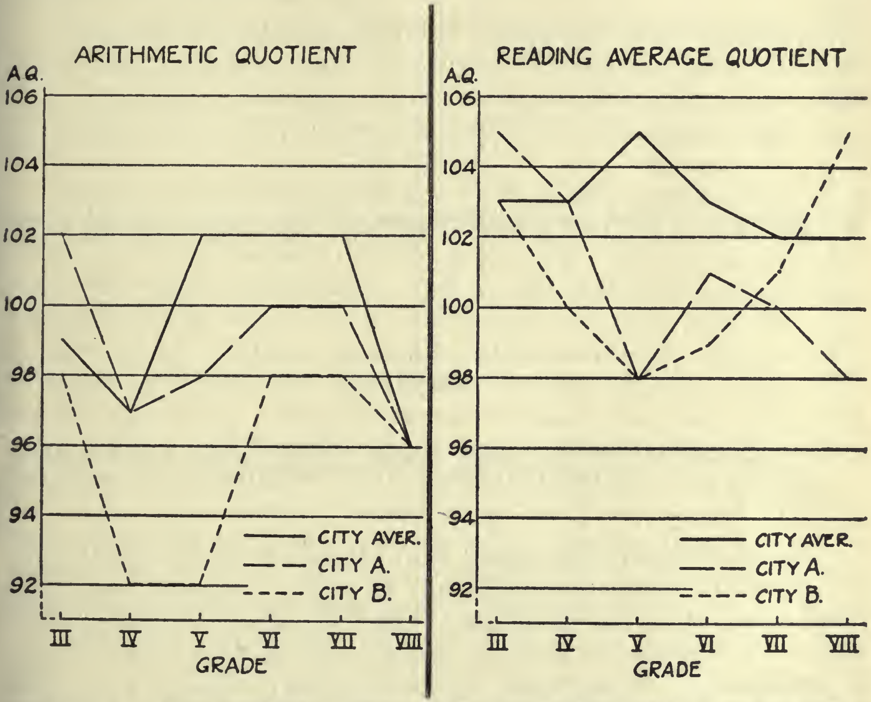


FIGURE 21. ACHIEVEMENT QUOTIENTS FOR EACH GRADE FOR CITY A AND CITY B.

A pupil's mental age constitutes an individual norm with which his achievement ages may be compared. The results of this comparison are expressed by the achievement quotients. The value of individual norms and the achievement quotient are illustrated by Figure 22. This figure shows the point scores and the achievement quotients for comprehension of silent reading of the pupils in a fifth-grade class. The former are plotted along the horizontal axis and the latter along the vertical axis. Distances of a dot from the two axes show the two measures of the pupil's achievement. The grade norm in terms of a point score is indicated by the arrow.

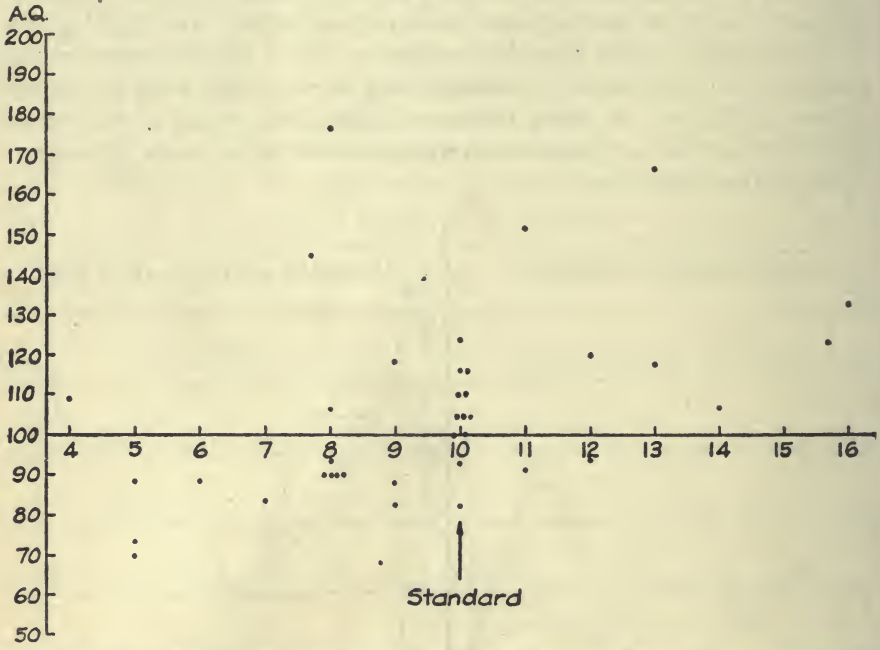


FIGURE 22. RELATION BETWEEN POINT SCORES AND ACHIEVEMENT QUOTIENTS IN COMPREHENSION OF SILENT READING, A FIFTH GRADE CLASS.

The lowest score in the class is four. If we had only the grade norm, all we could say about this pupil would be that he is conspicuously below standard and at the foot of his class. His A. Q. shows that in comparison with his own norm he has achieved more than is usually achieved by a pupil of his mentality. In fact when his mental age is considered, he is one of the "good" pupils in his class.

An added advantage of this plan is that when we have transmuted point scores in subject-matter tests into achievement ages, we have reduced them all to the same units in the sense that each successive year corresponds to an increment of ability gained by typical children in equal lengths of time. Since the achievement ages are expressed in the same units, they may be combined as point scores cannot be combined. For example, we have no obvious way of expressing the total achievement of a child who scores 155 in rate, 11 in comprehension, and 49 in arithmetic. By reference to Table XI, however, we observe that these point scores indicate, respectively, achievement ages of 12-6, 11-6, and 10-6 years. On the assumption that the abilities to which the point scores refer are all equally important, we may obtain the simple average of these ages, and we may thus express the composite achievement age as 11-6 years.

The translation of the point scores into age scores makes the measures more easily interpreted. Even without access to grade norms one is able to interpret partially a mental age of 13 or an achievement age of 9 years, 6 months. The quotients are also easily interpreted because in every case the norm is approximately 100.

The grade norms in the case of the achievement test do furnish appropriate educational objectives. The norms for the intelligence scale cannot, of course, be considered objectives because of the assumption that traits measured by this scale are not covered by school instruction.

VII. NORMS

For the Illinois Examination two types of norms are available: age norms and grade norms. The Illinois General Intelligence Scale has been standardized with respect to chronological age and the achievement tests with respect to mental age. Both the age scores and quotients have been standardized with respect to school grade. The grade norms are given in Table XVIII. The age norms are incorporated in the rules for translating the point scores into age scores. In both cases the norms are for *first trial* scores only. When the tests are repeated these norms should not be used. (See page 68). The grade norms for both ages and quotients are given in Table XVIII for three population groups; rural schools, city schools, and a general group, including city, town, and rural schools. The median scores for any grade or any city may be compared with these grade norms. It is helpful to make this comparison by expressing the median scores as deviations from these norms. When the median score is larger than the corresponding grade norm the deviation is positive, when less, negative. In Table XIX the median scores for a city are expressed as deviations from the grade norms. In the case of ages, the deviations are in terms of months: -6 indicates that the median age for this city is a half year below the grade norm, +8 that it is eight months above the grade norm.

TABLE XVIII. GRADE NORMS BASED UPON 55,000 OCTOBER SCORES*

Grade		Ages					Quotients				
					Silent Reading					Silent Reading	
		Mental	Arithmetic	Average	Comprehension	Rate	Intelligence	Arithmetic	Average	Comprehension	Rate
III	Rural	7–10	8–0	7–11	7–11	7–11	90	102	101	101	101
	Cities	7–11	7–10	8–2	8–2	8–2	89	99	103	103	103
	General	7–10	8–0	7–11	7–11	7–11	90	102	101	101	101
IV	Rural	9–4	9–3	9–1	9–0	9–1	96	99	97	96	97
	Cities	9–5	9–2	9–8	9–9	9–7	95	97	103	103	102
	General	9–4	9–2	9–5	9–4	9–5	95	98	101	100	101
V	Rural	10–6	10–11	10–1	10–0	10–2	98	104	96	95	97
	Cities	10–8	10–11	11–2	11–6	10–10	99	102	105	108	102
	General	10–8	10–11	10–8	10–9	10–7	98	102	100	101	99
VI	Rural	11–9	12–0	11–4	11–1	11–7	100	102	96	94	99
	Cities	12–0	12–3	12–4	12–6	12–2	101	102	103	104	101
	General	11–11	12–0	12–0	12–0	11–11	101	101	101	101	100
VII	Rural	13–1	13–2	13–1	13–1	13–0	101	100	100	100	100
	Cities	13–1	13–6	13–5	13–5	13–4	101	102	101	101	101
	General	13–1	13–4	13–4	13–3	13–2	101	101	101	100	101
VIII	Rural	14–3	14–3	14–2	13–10	14–6	106	103	100	97	102
	Cities	14–4	14–4	14–11	14–6	15–10	104	100	102	100	106
	General	14–3	14–2	14–7	14–2	14–10	104	101	101	98	104

*These are first trial scores.

TABLE XIX. MEDIAN SCORES FOR A CITY EXPRESSED AS DEVIATIONS FROM THE GRADE NORMS

Grade	Ages					Quotients				
	M.A.	Arith.	Aver.	Reading		I.Q.	Arith.	Aver.	Reading	
				Comp	Rate				Comp	Rate
III	—1	—2	—1	+4	—6	+2	—6	—1	+2	+0
IV	0	—3	0	—1	0	+2	—1	—2	+2	—4
V	—2	—6	—2	—3	0	+2	—5	+4	+4	+5
VI	+3	—9	—1	+8	—7	+6	—7	—1	—1	—2
VII	—2	—1	+4	+10	—1	0	+1	—2	0	—3
VIII	+4	+2	+1	+8	—7	0	+5	—1	+1	—3

The intelligence quotients derived from the Illinois General Intelligence Scale exhibit a greater degree of variability than those derived from the Binet Scale for Measuring Intelligence. For this reason one must use a different basis for interpreting the I. Q. 's in terms of degrees of brightness. Intelligence quotients derived from the Illinois General Intelligence Scale are estimated to be distributed as shown below* The appropriate interpretation is indicated in the left-hand column.

Degree of brightness	I.Q.	Percent of all children included
"Near" genius or genius	140 and above	1
Very superior	125–139	6
Superior	115–124	13
Normal or average	85–114	60
Dull	75–84	13
Border-line	60–74	6
Feeble-minded	Below 60	1

Achievement quotients exhibit a somewhat greater degree of variability than intelligence quotients. Their distribution differs from the normal by showing a greater degree of variability above the median (approximately 100) than below the median. It is, however, possible to divide the distribution so that the percent of pupils included in each division corresponds to that used for the interpretation of intelligence quotients. The following scheme for interpretation is suggested:

---

*It is necessary to estimate this distribution because the original rule for translating point scores into mental ages was found to be unsatisfactory.

Quality of pupils' achievement	Achievement Quotient	Percent of pupils included
Very superior	165 and above	1
	135–164	6
Superior	117–134	13
Average	83–116	60
Poor	71–82	13
	55–70	6
Failure	Below 55	1

Equivalence of Form 1 and Form 2. In considering norms for the two forms of the Illinois Examination it is necessary to inquire concerning the equivalence of the two forms. In case one form is easier than the other the same norms cannot be used. Data with reference to the equivalence of the scales which make up the Illinois Examination have already been given (See pages 9-18). These facts may be summarized and expressed in the following form which is more convenient to use. In order to reduce Form 2 age scores to the equivalence of Form 1 scores multiply by the following correction numbers:

	Correction Number
Illinois General Intelligence Scale	.98
Monroe's General Survey Scale in Arithmetic, Scale I.	.96
Monroe's General Survey Scale in Arithmetic, Scale II	1.00
Monroe's Standardized Silent Reading Test I, Revised, Comprehension	1.00
Rate	1.05
Monroe's Standardized Silent Reading Test II, Revised, Comprehension	.99
Rate	.97

Practise effect when a test is repeated. The grade norms given in Table XVIII are for the first application of the Illinois Examination. When it is given a second time pupils will tend to make higher scores because of their acquaintance with the nature of the tests. The amount of increase varies. If pupils are "coached" upon the tests a large increase is to be expected. When a period of several months intervenes between the first and second trials and the pupils have received no training upon the exercises of the tests, the increase appears to be small and in some cases can be neglected without serious error. No evidence was obtained relative to the effect of using the same form instead of a different form.

In order to ascertain the effect of practise when the second application immediately follows the first, both Form 1 and Form 2 were given to a number of pupils. Due to a miscarriage of plans the practise effect for the silent reading tests was not determined. For the Illinois General Intelligence Scale the average practise effect is approximately 5.0 points, or six months of mental age, if the returns from the eighth grade pupils are not used. In this grade unusual conditions appear to have prevailed and when the scores from it are included the practice effect is approximately 7.0 points. For Monroe's General Survey Scale in Arithmetic the average practise effect is approximately 3.2 points in grades III to V and 4.5 points in grades VI to VIII. It is, therefore, obvious that when the Illinois Examination is repeated after only a short interval the second trial scores must be corrected before they can be compared with those obtained from the first trial.

In one school the teachers of 134 pupils gave special drill and instruction to their pupils after Form 1 of the Illinois Examination had been given in November. The teachers did not know that Form 2 was to be given later, and did not have in mind, therefore, preparing the pupils for it. Their instruction was, to a slight extent, based upon Form 1 test papers. Believing that their pupils were rather weak in knowledge of vocabulary and in synonym-antonym, special drill was given along these lines in language work. In arithmetic practise was given upon those combinations where the pupils seemed weak. In reading there was some special drill for increasing the rate of silent reading. The gains made by the pupils under these teachers are given in Table XX.

It is commonly assumed that general intelligence is unaffected by school instruction. A median gain of 43.2 points or slightly more than four years in mental age within a period of six months indicates that mental age, as measured by such an instrument as the Illinois Intelligence Scale, is affected by class room instruction. It is, therefore, necessary to exercise a great deal of caution in interpreting changes in the intelligence scores derived from two successive testings separated by a considerable time interval. Even in the case of a first trial the scores obtained will be misleading if the pupils have received any special preparation for the test.

One would naturally expect that the achievement scores in arithmetic and silent reading would be materially affected by instruction. Table XX shows very large gains in achievement. The increases in achievement quotients show that except for the comprehension of silent reading the gains are relatively larger in achievement than in mental age. It is not unlikely that some of the increases in the May scores over the November scores are due to the pupils being more familiar with the testing procedure. The effect of familiarity was not investigated but due caution

TABLE XX. GAINS DUE TO SPECIAL INSTRUCTION UPON THE ILLINOIS EXAMINATION

	Median Point Scores			Median Quotients		
	Nov. 1920	May 1921	Gain	Nov. 1920	May 1921	Gain
Intelligence.....	57.6	100.8	43.2	99.6	128.0	28.4
Arithmetic......	48.2	120.0	71.8	114.0	137.4	23.4
Comprehension..	10.9	15.5	4.6	105.1	102.0	—3.1
Rate...........	171.4	237.9	66.5	122.0	139.4	17.4

should be exercised in interpreting the gains in achievement as being the result of instruction directed to the needs of the pupils.

UNIVERSITY OF ILLINOIS BULLETIN

ISSUED WEEKLY

VOL. XIX DECEMBER 12, 1921 No. 15

[Entered as second-class matter December 11, 1912, at the post office at Urbana, Illinois, under the Act of August 24, 1912. Acceptance for mailing at the special rate of postage provided for in section 1103, Act of October 3, 1917, authorized July 31, 1918.]

BULLETIN NO. 7

BUREAU OF EDUCATIONAL RESEARCH, COLLEGE OF EDUCATION,
UNIVERSITY OF ILLINOIS

Types of Learning Required of Pupils in the Seventh and Eighth Grades and in the High School

BY

WALTER S. MONROE

Director of Bureau of Educational Research

LIBRARY
SEP 29 1922
UNIVERSITY OF TORONTO

PRICE 15 CENTS

PUBLISHED BY THE UNIVERSITY OF ILLINOIS
URBANA, ILLINOIS

INTRODUCTORY STATEMENT

The study reported in this bulletin is a sub-project of a more extended investigation of the study habits of pupils. This study was made partly to secure information relative to the types of learning required of pupils and partly for the purpose of bringing certain characteristics of the learning process to the attention of teachers. This report is based upon the information secured by means of a questionnaire. This method of investigation was followed because it furnished the only feasible plan of inquiring into current practices.

A large number of teachers contributed to this study. The writer wishes to express his appreciation of their cooperation and submits this report in the hope that it will reward them for the time and effort invested in the project.

WALTER S. MONROE, *Director.*

Types of Learning Required of Pupils in the Seventh and Eighth Grades and in the High School

Nature of the learning process. Our plan of education requires that the pupil in the upper grades of the elementary school and in the high school spend a large share of his time in studying textbooks. In making an assignment, the teacher directs the pupil to study certain pages in the basic text and, in some cases, to supplement this by the study of certain references. In the preparation of these assignments the pupil learns. In considering the nature of the learning process, it is necessary to recognize that all assignments do not call for the same kind of mental activity. This is true even when we limit our consideration to assignments that call for study in which silent reading is the central activity. For example, one type of mental activity is required in memorizing a poem, another type in studying a theorem in geometry. The preparation of a lesson in history requires a different mental activity.

The type of learning should be specified in the assignment. A good assignment should indicate to the pupil the kind of study in which he is expected to engage as well as the pages of the textbook which he is to study. A good assignment includes a number of other things, but the pupil should not be left to his own resources to determine the procedure which he shall follow in the preparation of his lesson. It is, therefore, pertinent that we inquire concerning the types of study or learning required of pupils in the various subjects taught in the seventh and eighth grades and in the high school.

The problem of this investigation. The problem of this investigation is to ascertain, by means of a questionnaire submitted to teachers, the major types of textbook study required of pupils in the following school subjects: (1) geography, (2) history and civics, (3) literature, (4) physics, (5) chemistry, (6) physiology, (7) biology, (8) agriculture, (9) arithmetic, (10) algebra, (11) geometry.

Source of data. On September 29, 1921, a circular letter was addressed to the city superintendents and principals of high schools in Illinois, asking them to bring to the attention of their teachers the general plans of "an investigation of the study habits of high school pupils." They were asked to send to the Director of the Bureau of Educational Research the names of those teachers in the seventh and eighth grades and in the high school who were willing to cooperate in this investigation. The questionnaire, on which this report is based, was sent only to those teachers whose names were secured in this way. About 800 copies were sent out. At the time this report was written only 317 questionnaires had been returned. In several cases a note accompanied the blank, saying that it represented the consensus of opinion of several teachers. Consequently, this report may be thought of as representing the consensus of opinion of more than 317 teachers. The teachers to whom the questionnaires were addressed were directed to answer the questions for only the school subjects which they were then teaching or had taught.

Basis of analysis of types of learning. In order to provide a uniform basis upon which teachers might state the types of learning required in these school subjects, it was necessary to analyze the learning process and to identify the most significant types of learning. This analysis was based on the products or outcomes of the learning process. Learning is a productive activity. As a result of engaging in study the pupil produces something. This may be simply ideas stored up in his mind or it may be a written record, such as a summary or an outline. We may assume that to a classification of outcomes there is a corresponding classification of types of learning. For example, one outcome of study is the production of a summary of the material studied. This requires a definite type of learning, It is entirely different from the type of learning required when the outcome is the memorization of the material studied.

The analysis presented in the following pages recognizes twelve outcomes or types of learning. The analysis was intentionally restricted to textbook study in which reading is the central activity. This eliminates most of the study in the field of spelling and the learning which the pupil does in connection with working examples in arithmetic or algebra. This limitation was made not because the other types of learning are not important but because it was felt desirable to confine this investigation to the types of learning in which reading is the central activity.

Furthermore, indirect or generalized outcomes of study were not included. These outcomes are not the products of studying a particular assignment but are, rather, built up over a long period of study and their realization occurs as a by-product of the production of direct or specific outcomes. Included in these generalized outcomes are:

1. Acquisition of more effective methods of thinking and reasoning.
2. Interest in and appreciation for a given field of study.
3. Acquisition of more effective modes of expression.

Types of learning in textbook study. In the list of outcomes of textbook study given below, there is some overlapping. Furthermore, it must be recognized that two or more of these outcomes may be produced simultaneously. However, it has seemed wise to recognize the distinctions implied by this list for the purpose of analyzing the process of study.

I. Comprehension of material read plus memorization so that it can be reproduced.

> This reproduction need not be verbatim. It may be given in the pupil's own words, but it should include most of the minor ideas as well as the central ones. This reproduction may be given in the form of "free composition," as in the case when the pupil is simply directed *to tell* what he has read. As a special case, the reproduction may take the form of "answer composition" in which the pupil is asked specific questions which he answers in terms of what he has studied.

II. With the text at hand preparation of a summary which contains the central ideas of the assignment studied.

> In this summary the central ideas should be given but the supporting details are not essential. Neither is it expected that the relative importance of the central ideas be indicated. It is not expected that this summary will be given in outline form, which is recognized as a different outcome.

III. With the text at hand preparation of an outline which gives the principal points and supporting details arranged to show order of relative importance and relations to each other.

> This outcome overlaps the preparation of a summary. An outline may be considered a summary but a summary is not an outline. There are also differences other than those of form.

IV. Obtaining information for the purpose of solving problems, or answering questions.

> In the production of this outcome the pupil has the ques-

tions or problems before him as he studies. He is getting information for specific purposes. The problems or questions with which he is dealing may vary widely with reference to scope. The required answer may be a single fact or it may involve the preparation of a well organized report covering several pages.

V. Extension of one's range of general information by reading widely material directly related to a given subject.

This outcome is general and indefinite. It is difficult to define in explicit terms. The information which the pupil acquires may not be very precise and so far as this outcome is concerned it need not be well organized, although it relates to some general topic.

VI. Discovery of collateral or illustrative material for topics or problems under discussion.

This type of outcome is closely related to and overlaps somewhat with V. It is, however, believed that a distinction is worth while. The pupil's purpose is somewhat different when he is looking for collateral or illustrative material rather than information to use in the solving of a problem.

VII. Enlargement of vocabulary.

Vocabulary may be enlarged in two ways. One may become acquainted in a general way with new words. Such acquaintance with words is sufficient for a great deal of general reading. Another way in which a vocabulary may be enlarged is by the determination of the technical or precise meaning of the words peculiar to a given subject. In most school subjects there are some words which have a meaning which is peculiar to that subject. These words may occur in general reading or in other subjects with a different meaning or they may be words which are found only in the field of the one subject. This outcome overlaps a number of the other outcomes enumerated here; or perhaps it is better to say that it is produced simultaneously with other outcomes. It is, however, possible to think of pupils studying primarily for this purpose.

VIII. Appreciation of the significance of each word used in a concisely expressed statement or principle.

This type of outcome results from the intensive study of an assignment with reference to the implications of all words contained within it in richness of meaning or emphasis.

IX. A clear comprehension of the essential conditions of a problem which is to be solved.

This is essentially understanding what is given and what is to be found or done in the solving of a problem.

X. Discovery of new or supplementary problems related to the topic being studied.

In the production of this outcome the pupil is maintaining a critical or questioning attitude. The discovery of new or supplementary problems may grow out of a comparison of a statement by one author with a statement by another or out of the comparison of the assignment with the pupil's experience.

XI. Drawing valid conclusions from given data or statements.

A special case of this occurs in the determination of the validity of statements or inferences. In this case the given statements or inferences are compared with other statements or with one's general store of information and a conclusion is drawn. This is essentially nothing more than drawing valid conclusions from the data or statements at hand.

XII. Following directions with accuracy and reasonable speed.

This outcome is produced only when the assignment to be studied consists of directions to be followed.

The questionnaire. A copy of the exposition of the twelve types of learning which we have just given accompanied the questionnaire. The teachers were directed to become familiar with the twelve types of learning before they attempted to answer the questions. The questionnaire is reproduced on page 8:

BUREAU OF EDUCATIONAL RESEARCH—QUESTIONNAIRE NO. I
TYPES OF TEXTBOOK STUDY

Name________________________________ Address______________________________

School Subject__________________________ School Grade___________________

Below we list a number of school subjects in which textbook study is required. Several school subjects are not included because they appear to require peculiar types of study. Answer the following questions *only for the subjects which you are now teaching or have taught.* If you have never taught any of the subjects included in the list below write in the name of a subject which you have taught and answer the questions with reference to it.

1. What are the *major* types of textbook study which you require of your pupils in the following subjects?

Answer this question in terms of the outcomes enumerated above. Designate the types of study by means of the numbers assigned to the outcomes above. Give these numbers in what you consider to be the order of importance of the types of textbook study for this school subject. Give the most important first. It is recognized that two or more types of textbook study may be required in a single subject:

SUBJECT	TYPES OF STUDY	SUBJECT	TYPES OF STUDY
Geography .	______________________	Biology	______________________
History and Civics	______________________	Agriculture	______________________
Literature ..	______________________	Arithmetic .	______________________
Physics	______________________	Algebra ...	______________________
Chemistry ..	______________________	Geometry ..	______________________
Physiology .	______________________		______________________

NOTE: Biology is intended to include botany and zoology even when they are taught as separate subjects. All sciences are intended to include the study of both the text and the laboratory manual in case one is used. In the case of arithmetic, algebra, and geometry it is our intention to include only those phases of the subject which require the study of the text.

2. Are there other significant outcomes of textbook study in which reading is the central activity that should be added to the list of twelve which we have given?____________________If so, what are they?____________________

____

3. In your judgment which of the twelve types of study do pupils find most difficult? (Give your answers in order of importance, giving the most difficult first.) ____

____

____

4. What types of study offer the greatest difficulty to the teacher in training pupils in methods of study?____

____

____

Limitations of replies to questionnaire. In interpreting the summaries of the replies to the questionnaire certain limitations must be kept in mind. In the first place, the analysis of the learning process and the recognition of different types of learning was probably unfamiliar to many, if not to all, of the teachers who answered the questionnaire. The brief exposition of this analysis and the description of the types of learning were, doubtless, not sufficient in many cases to give the teachers an adequate understanding of the twelve types of learning. From letters which occasionally accompanied the questionnaires and from information secured in other ways, it appears that the questions were answered with a great deal of care. In several instances, where a number of teachers from the same school were cooperating, they met together and discussed the questions before answering them.

The questions were asked with reference to present practice in the schools. However, the answers must be thought of merely as opinions of the teachers; for the present practice is, doubtless, a reflection of their judgment concerning what should be asked of their pupils. The summaries presented in the following tables should, therefore, be thought of as representing the concensus of the opinions of teachers in regard to present practice.

Types of learning required by pupils in certain school subjects. The replies to the first question are summarized in Table I, which is to be read as follows: Reports were secured from 65 teachers of geography. Of this number, 56 mentioned Type I as a major type of textbook study required in geography; 27 mentioned Type II; 37 mentioned Type III; and so forth. Table II is derived from Table I by calculating the percent of teachers who mentioned each type of study. These percents are given only for the subjects appearing on the questionnaire blank. It appears that teachers had considerable difficulty in defining major types of textbook study. Some were inclined to mention practically all types of textbook study as being major types while others restricted their answers to one or two types only. The frequencies of the number of major types mentioned are summarized in Table III. Of the 65 teachers reporting for geography, only one restricted his answer to one major type; seven mentioned two major types; 24 mentioned three major types; and so on. The number of types most frequently mentioned by a teacher is three. Four teachers mentioned all twelve types.

TABLE I. FREQUENCY OF MENTION OF TYPES OF STUDY IN VARIOUS SCHOOL SUBJECTS

SUBJECTS	Number of Teachers	TYPES OF STUDY												TOTAL
		I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	
Geography	65	56	27	31	35	37	33	10		7	11	16	13	276
History and Civics	103	80	45	57	44	65	45	16	4	14	25	32	9	436
Literature	100	88	53	47	34	50	23	87	50	6	14	18	17	487
Physics	47	30	8	9	30	9	11	7	13	22	14	36	25	214
Chemistry	34	24	8	7	22	10	6	6	8	13	13	24	21	162
Physiology	49	40	17	11	17	12	9	16	2	6	12	14	14	170
Biology	44	34	15	10	21	20	15	13	3	5	12	17	20	185
Agriculture	29	17	7	8	22	10	9		3	8	11	13	9	117
Arithmetic	65	12	3	3	40	1	4	8	14	56	26	22	36	225
Algebra	66	20	3		44	1	5	7	25	58	21	31	40	225
Geometry	60	25	3	1	33		2	11	20	50	26	45	34	250
Industrial Education	4	1			2	1				1	1		1	7
Bookkeeping and Shorthand	4	1	1		2			2	2	3	1	2	4	18
Manual Training	4	1			2	2	1	2		2			2	12
Latin	17	14	2	2	8	3	1	14	7	4	4	5	6	70
Music	1	1						1	1				1	4
Grammar, Rhetoric, and Composition	8	5	1	1	1		1	1	1	1			3	15
General Science	6	3			4	1	2	1	1	2	2	6	5	27
Spelling	1							1						1
French	4	3	3		1	4		3	1					15
Household Arts	13	10	6	3	9	4	4	6	1	7	4	5	12	71
Penmanship	2												2	2
Trigonometry	1	1			1					1		1	1	5
Economics	3	1		2	2	3		1		2	3	1		15
Total	729	467	202	192	377	233	171	213	156	268	200	288	275	3042

TABLE II. PER CENT OF TEACHERS MENTIONING EACH TYPE OF STUDY

SUBJECTS	Number of Teachers	TYPES OF STUDY											
		I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
Geography	65	86	42	48	54	57	51	15		11	17	25	20
History and Civics	103	78	44	55	43	63	44	16	4	14	24	31	9
Literature	100	88	53	47	34	50	23	87	50	6	14	18	17
Physics	47	64	17	19	64	19	23	15	28	47	30	77	53
Chemistry	32	75	25	22	66	31	19	19	25	41	41	75	66
Physiology	48	83	35	23	35	25	19	33	4	13	25	29	29
Biology	44	77	34	23	48	45	34	30	7	11	27	39	45
Agriculture	27	63	26	30	81	37	33		11	30	41	48	33
Arithmetic	67	18	4	4	60	1	5	12	21	84	39	33	54
Algebra	68	30	4		65	1	7	10	37	85	31	46	59
Geometry	61	41	5	2	54		3	18	33	82	43	73	56

TABLE III. NUMBER OF MAJOR TYPES OF TEXTBOOK STUDY

SUBJECTS	NUMBER OF MAJOR TYPES MENTIONED												Number of Teachers
	1	2	3	4	5	6	7	8	9	10	11	12	
Geography	1	7	24	13	7	7	3		1	2			65
History and Civics	2	11	20	22	22	13	3	2	4	1	1	2	103
Literature	3	3	22	17	21	20	4	5	2		2	1	100
Physics		5	14	9	6	5	4	2	1	1			47
Chemistry		2	6	6	8	2	3	3	1	1			32
Physiology	5	9	15	4	7	5	1	2					48
Biology	1	4	10	12	5	7	5						44
Agriculture	2	3	8	7	2		2		2	1			27
Arithmetic	8	10	20	15	6	3	4	1					67
Algebra	4	11	20	13	7	5	8						68
Geometry	1	10	14	13	7	7	7	1		1			61
Industrial Education		2	1		1								4
Bookkeeping and Shorthand	2			1	1								4
Manual Training	1	1	1	1									4
Latin	1	3	4	2	3	2	1					1	17
Music				1									1
Grammar and Rhetoric	1	2		2	1	1							7
General Science		1		3		1	1						6
Spelling	1												1
French			1			3							4
Household Arts		2	2	1	1	3	2	1	1				13
Penmanship	2												2
Trigonometry					1								1
Economics			1	1				1					3
Total	35	86	183	143	106	84	48	18	12	7	3	4	729

In interpreting Tables I and II, it is necessary to keep in mind the above limitations. These tables give a summary of the opinions of teachers concerning the types of study which are most important in the several subjects. Since these types of textbook study were defined in terms of outcomes, it follows that we have, also, an indication of the types of outcomes which teachers consider most important. It would be profitable to compare these outcomes with statements of educational objectives. It is not unlikely that, if teachers were asked concerning the outcomes which they considered most important in these subjects, a different consensus of opinion would be secured. The following observations may be made with reference to these three tables. These statements may be thought of as tentative conclusions:

1. In the study of a school subject a pupil is required to engage in several types of learning. In other words, in a given school subject, study generally means the doing of several things rather than of a single thing.

2. Table III indicates a variety of opinion concerning the number of major types of study required of pupils in the different subjects. Some teachers appear to be restricting the learning process to one or two types, while others are requiring many types of learning.

3. The most frequent types of learning are: I, ''comprehension of material read plus memorization so that it can be reproduced''; IV, ''obtaining information for the purpose of solving problems or answering questions''; XI, ''drawing valid conclusions from given data or statements''; IX, ''a clear comprehension of the essential conditions of a problem which is to be solved''. The least frequently mentioned is VIII, ''appreciation of the significance of each word used in a concisely expressed statement or principle''.

4. No type of study is mentioned by 100 percent of the teachers. This fact, together with the fact that all types of learning were reported for a number of school subjects, indicates that some teachers are not in agreement with the majority with respect to the types of learning required of their pupils.

5. The most frequently mentioned types of learning vary for the different subjects. For example, in Literature, 88 percent mention Type I and 87 percent Type VII. In Physics, 77 percent mention Type XI and 64 percent Types I and IV. In Arithmetic, the most frequently mentioned is Type IX, which was given by 84 percent. Type IV was mentioned by 60 percent and Type XII by 54 percent.

TABLE IV. RELATIVE DIFFICULTY OF TYPES OF LEARNING FROM THE STANDPOINT OF THE PUPIL

	TYPES OF LEARNING												
RANK	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	TOTAL
1	44	14	37	19	9	14	4	21	39	24	55	16	296
2	8	22	21	24	14	19	9	39	35	29	40	20	280
3	11	7	15	20	7	13	10	19	39	22	42	28	233
4	5	5	13	12	4	7	7	16	16	14	20	21	140
5	5	1	5	1	1	13	4	9	4	9	13	8	73
6	1	3	1	7	3	2	1	6	1	4	3	6	38
7	2		1	2	5	3	5	1	1	2	2	4	28
8		2	2	5	4	2	2		1		2	2	22
9	4	1		4	1	2	3	3				1	19
10	1	4	5	1	2	1	2		1			1	18
11	4	5	1		3	1	2	1		1			18
12	3	1		1	1		3	1	1	2		4	17

TABLE V. RELATIVE DIFFICULTY OF TYPES OF LEARNING FROM THE STANDPOINT OF INSTRUCTION

TYPES OF LEARNING												
I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	TOTAL
63	48	67	71	60	63	51	87	120	95	130	87	942

6. The fact that different subjects demand different types of learning makes it necessary that the teacher of each subject give training to the pupils in the type of learning required by that subject.

The relative difficulty of types of learning from the standpoint of the pupil. The answers to Question 3 have been summarized in Table IV. Only 296 replies to this question were received. Type I was mentioned by 44 teachers as being the most difficult type of learning from the standpoint of the pupil. Type II was mentioned by 14 teachers; Type III, by 37 teachers; and so on. The totals in the column at the right of the table show that 16 teachers mentioned only one type of learning (296-280). Only 73 teachers mentioned five or more types. Out of 296, 17 mentioned all twelve types. The marked diversity of opinion with reference to the relative difficulty of these twelve types of learning is probably due, in part, to the fact that these teachers were answering with reference to different school subjects. It was also doubtless due, in part, to the fact that the teachers were not familiar with this analysis of the learning process. However, this table should give some indication of the relative difficulty of these types of learning. It is interesting to note that Type I, ''comprehension of material read plus memorization'', is indicated as being the second most difficult type of learning. It is exceeded in difficulty, according to the judgment of these teachers, only by Type XI, ''drawing valid conclusions from given data or statements''. Type V, ''extension of one's range of general information by reading widely material directly related to a given subject'', is indicated as being the least difficult.

The relative difficulty of types of learning with reference to instruction. The answers to Question 4 are summarized in Table V. Question 4 asked concerning the difficulty of training pupils in these twelve types of learning. The teachers were asked to indicate the types of study which ''offer the greatest difficulty to the teacher in training pupils in methods of study''. Most teachers mentioned two or more types of study. In Table V, no distinction is made with reference to the relative difficulty of the types mentioned by the teachers. Type XI, ''drawing valid conclusions from given data or statements'', is judged to be most difficult. It is also considered to be most difficult from the standpoint of the pupil.

Additional types of learning suggested. Question 2 asked the teachers to indicate any significant outcomes of textbook study that should be added to the list which defines the twelve types of learning. Only 68 teachers suggested additional outcomes. Most of the suggestions were with reference to generalized outcomes, which we excluded from our analysis. Development of character, training in concentration, and development of appreciation for a given field of study were mentioned as additional generalized outcomes.

A number of outcomes of study not based upon silent reading were also mentioned. This investigation was definitely restricted to the types of study in which reading is the central activity. Therefore, these suggestions do not apply to our analysis, although they do pertain to important types of study.

''The organization of ideas gained from study with results of other study and with ideas gained from experience'' was mentioned by several teachers. This is, undoubtedly, an outcome of study and defines an important type of study. Somewhat the same idea was suggested by a few teachers under the head of ''application of ideas and principles''. ''Participation in and understanding of the experiences of others'' was mentioned as an outcome of study by one teacher. Another teacher mentioned ''making out questions based upon assignments studied''. These suggestions for additional types of study, even though relatively few, indicate that some modifications should, perhaps, be made of the analysis presented to the teachers.

BULLETIN NO. 8
BUREAU OF EDUCATIONAL RESEARCH
COLLEGE OF EDUCATION

A CRITICAL STUDY OF CERTAIN SILENT READING TESTS

By
WALTER S. MONROE, DIRECTOR

PUBLISHED BY THE UNIVERSITY OF ILLINOIS
URBANA

TABLE OF CONTENTS

PAGE
Preface 4
The Measurement of Silent Reading Ability 5
The Problem 5
The Data Collected. 6
The Performances Required of a Pupil 9
Description of Pupils' Performances 13
Scoring Reproductions 15
The Idea-Counting Method 16
Brown's Method of Idea Counting 17
The Word-Counting Method 19
Subjectivity of Describing Reproductions 20
Constant Errors and Variable Errors 20
Summary for Describing Reproductions 25
Scoring Answers to Questions 25
Describing the Quality of Compositions 25
Time Required for Scoring Test Papers 25
Average Scores and Standard Deviations 26
Equivalence of Duplicate Forms 29
Relation of Vocabulary to Difficulty 30
Formation of Composite Scores 31
Reliability 32
Methods of Determining Reliability 32
Probable Error of r Due to Sampling 33
Reliability of Tests Studied 34
Discrimination 37
Comparison with Teachers' Ratings 39
Correlation of Comprehension with Memory 39
Corrected Coefficients of Correlation 41
Correlation of Comprehension with Vocabulary 42
Correlation of Cancellation Scores with Measures of Rate of Reading 44
Correlation of Comprehension with Written Composition 44
Inter-correlation between Tests 46
Correlation of Single Tests with Composites 50
Summary of Conclusions 51
Correlation with Composites 52

PREFACE

In the field of silent reading, as well as in the fields of other school subjects, the number of available educational tests has been increased so that one desiring to use a test is confronted with the necessity of making a choice. If such a choice is to be made intelligently it is necessary to have at hand experimental data with reference to the reliability and validity of the tests considered. The study which is reported in this monograph was undertaken for the purpose of securing such data with reference to certain silent reading tests. The report is presented in hopes that users of silent reading tests will find the information that it contains helpful in making an intelligent selection of educational tests in this field. The monograph will doubtless also be of interest to students in the field of educational measurements.

Walter S. Monroe,
Director, Bureau of Educational Research.

A CRITICAL STUDY OF CERTAIN SILENT READING TESTS

The measurement of silent reading ability. The scores yielded by silent reading tests may fail to be true measures of silent reading ability for two reasons. *First,* the scores may not be reliable or accurate. A score is lacking in reliability when two applications of a test or of duplicate forms of it do not yield approximately the same score when administered to the same pupils, as far as possible, under the same conditions. Included in this is any lack of objectivity in the scoring of the test. *Second,* the performance which a pupil gives on a silent reading test may depend upon other factors in such a way that it is an index of these factors rather than of silent reading ability. For example, when a pupil answers questions from memory his answers may be influenced to such an extent by his ability to remember that his performance is not a truthful index of his ability to read silently.

Two aspects of the activity of silent reading may be recognized. *First,* the reading mechanism consists of perception, eye-movement habits, etc. The rate of silent reading is largely dependent upon this mechanism and hence any measure of rate is an index or symptom of the quality of the mechanism. *Second,* the thought-getting or comprehension aspect of silent reading involves the higher mental processes. The quality of this is indicated by the comprehension scores. Comprehension is not entirely independent of the mechanism of silent reading, but, if sufficient time is allowed, pupils who possess poor reading mechanism may stand high in thought-getting.

The problem. The problem of this study is to ascertain the reliability and, so far as possible, the function and validity of certain silent reading tests. These tests, as will be shown later, differ in the performances which are required of the pupils. They also differ in other respects. Their titles suggest that all of the silent reading tests included in this study are designed to measure silent reading ability. The fact that they differ widely in certain respects suggests the possibility that no two of them measure the same type of reading ability, or at least that they do this with different degrees of validity. The study has been restricted to tests which yield some measure of the rate of reading as well as a measure of comprehension in order that the measurement of both phases of silent reading activity might be studied. With one exception, the tests which have been used have duplicate forms. In addition to the silent reading tests, certain other tests were given to the same pupils, because it was thought that the scores yielded by them might assist in the analysis and interpretation of the scores yielded by the silent reading tests.

The data collected. Through the courtesy of Superintendent W. W. Earnest and certain teachers of the Champaign Public Schools, the tests chosen for this study were given in the spring of 1920 to a number of pupils in the fourth and seventh grades. All of the tests were administered by Miss Dora Keen, at that time a research assistant in the Bureau of Educational Research. Care was exercised to secure as nearly uniform testing conditions as can be obtained in the ordinary schoolroom. The lapse of time between the giving of the different forms of the same test was made as nearly equal as possible for the different groups. Only in rare instances were tests given after recess in the afternoon or during the afternoon session on Friday. The tests were given to all pupils in four rooms in both the fourth and seventh grades. The total number of pupils tested in each grade was approximately 140. The study is, however, based upon the records of only those pupils who took all of the tests. The number of complete records in the fourth grade is 80 and in the seventh grade, 91.

The following tests were given in the fourth grade:

1. The Courtis Silent Reading Test No. 2[1], Form 1, "The Kitten Who Played May Queen," and Form 3, "The Kitten Who Caught a Fish."

2. Brown's Silent Reading Test[2], Form 1, "The Long Slide," and Form 2, "A Morning Adventure."

3. Monroe's Standardized Silent Reading Test I[3], Forms 1, 2, and 3.

---

[1] Courtis Silent Reading Test No. 2. Forty-sixth Annual Report. Kansas City, Missouri: Board of Education, 1917. pp. 79-85.

[2] Brown, H. A. "The Measurement of Ability to Read." A Manual of Directions Concerning Giving and Scoring of Reading Tests, Statistical Treatment of the Data and Diagnosis of School Class and Individual Needs. Concord: New Hampshire Department of Public Instruction (in cooperation with the General Education Board). Bureau of Research Bulletin No. 1, Second Edition, 1916. pp. 57.

[3] Monroe, W. S. "Monroe's Standardized Silent Reading Tests." Journal of Educational Psychology, 9:303-12, June, 1918.

4. Fordyce's Scale for Measuring Achievement[4] in reading Test No. 1, "Narcissus."

5. Experimental Reproduction Test I, Form 1, based on pages 84 and 85 of the supplementary reader, "The Strike at Shane's"[5], and Form 2 based on pages 6 and 7 of the same publication. The passage for Form 1 contains 370 words and that for Form 2, 395 words. In administering these tests the pupils read from the supplementary reader. The exact place of beginning had been marked in each copy. Also the end of the passage to be read was indicated.

6. Cross-Out Silent Reading Test I, Form 1 and Form 2. This is an experimental silent reading test. In a passage of rather simple reading material, words were substituted, which did not agree with the meaning of the preceding words in the sentence. A pupil is asked to cross out the words which do not fit. With the exception of the substituted words, the selection is a connected story.

7. Vocabulary Test. The words of this test are those used by Terman and Childs. The form of the test is that proposed by Whipple[6].

8. Cancellation Test, "a-t" and "e-r[7].

9. Memory, "How Mr. Lincoln Helped the Pig."[8]

The following tests were given in the seventh grade:

1. Starch's Silent Reading Test No. 6 and Test No. 7.[9]

2. Monroe's Standardized Silent Reading Test II, Forms 1, 2, and 3.

3. Fordyce's Scale for Measuring Achievement in Reading, Test No. 2, "Spirit of Spring."

---

[4] Fordyce, Charles. "A Scale for Measuring the Achievements in Reading." The University Publishing Company, Lincoln, Nebraska, and Chicago. 1916.

[5] "The Strike at Shane's." (Gold Mine Series, No. 2.) Boston: American Humane Education Society, 1908. pp. 91.

(A supplementary reader for the fourth grade which has as its lesson kindness to domestic animals.)

[6] Whipple, G. M. Manual of Mental and Physical Tests, Complex Processes Chapter 12. Baltimore: Warwick and York, 1914.

[7] This test is described by Whipple in the Manual of Mental and Physical Tests, Simpler Processes, p. 311.

[8] Whipple, G. M. Manual of Mental and Physical Tests, Simpler Processes, Pages 207-10.

[9] Starch, Daniel. The Measurement of Efficiency in Reading. Journal of Educational Psychology, 4:1-24, 1915. These tests were used as duplicate forms.

4. Experimental Reproduction Test II, Form 1, based on pages 6 and 7 of the supplementary reader, "Old English Heroes,"[10] and Form 2, based upon pages 8 and 9 of the same publication. The passage for Form 1 contains 662 words, and that for Form 2, 611 words.

5. Cross-Out Silent Reading Test II, Form 1 and Form 2. This test is similar to the Cross-Out Silent Reading Test used in the fourth grade but is based upon more difficult material.

6. Pressey Silent Reading Test for Grades VI, VII, and VIII, Form 1 and Form 2. This is an experimental test.

7. Vocabulary Test. This is the same test as that used in the fourth grade.

8. Cancellation Test, "a-t" and "e-r." This is also the same as that used in the fourth grade.

9. Memory Test, "Marble Statue."[11]

10. Composition Test. The Willing Composition Scale[12] and the directions which accompany it were used.

In addition to the above tests a rating for ability in silent reading was secured from the teachers. To guide them in making this rating, the teachers were given the following directions:

> Think of all the fourth (seventh) grade pupils with whose silent reading ability you have ever become acquainted from the best to the poorest. Compare each child in your present class with this distribution of pupils. Give a pupil a rating of 5 if he has *very superior* ability in silent reading equalled only by about seven out of every hundred, or 7 percent of fourth (seventh) grade pupils. Give him a rating of 4 if he has *superior* ability or ability above the average, yet is excelled by the *very superior* group. About 24 out of every hundred, or 24 percent of fourth (seventh) grade pupils, will fall in the *superior* group. Give him a rating of 3 if he possesses average ability, i. e., ability which lies somewhere close to the middle of the difference between the very best pupil and the very poorest. About 38 out of every hundred, or 38 percent of fourth (seventh) grade pupils, will fall in this average group. If the pupil is below the average in ability to read and yet

---

[10] Bush, Bertha E. Old English Heroes. (Instructor Literature Series—No. 116.) Danville, N. Y., and Chicago: F. A. Owen Publishing Co., and Hall and McCreary, 1909. Pp. 31.

This is a supplementary reader suitable for the upper elementary grades. It contains brief sketches of the lives of Alfred the Great, Richard the Lion-Hearted, and the Black Prince.

[11] Whipple, G. M. Manual of Mental and Physical Tests, Simpler Processes, Pages 107-10.

[12] Willing, M. H. Measurement of Written Composition in Grades IV to VIII, English Journal, 7:193-202, March, 1918.

does not equal the poorest you have ever known give him a rating of 2. This group is called *inferior* and will contain about 24 out of every hundred, or 24 percent of fourth (seventh) grade pupils. Give the pupil a rating of 1 if he is *very inferior* in ability to read so that he is as poor or very nearly as poor as the poorest pupil you have ever known. About 7 out of every hundred, or 7 percent of fourth (seventh) grade pupils, will fall in this very inferior group.

The above directions do not mean that you will necessarily be obliged to give 7 percent of your class a rating of 5; 24 percent, a rating of 4; 38 percent, a rating of 3; 24 percent, a rating of 2; and 7 percent, a rating of 1. They do mean, however, that a large number of pupils, a number running up into the hundreds, can be divided in exactly this manner, i. e., 7 percent, very superior; 24 percent, superior; 38 percent, average; 24 percent, inferior; and 7 percent, very inferior. You are to think of all the pupils you have ever known from the best to the poorest and by comparison give each pupil in your present class the rating he would receive if he were included with all the pupils you have known and the entire number should be rated in the above manner.

The performances required of a pupil. All of the silent reading tests in the above list are designed to measure the ability to read silently. However, they require a variety of performances from the pupil. In the Courtis Silent Reading Test No. 2, the pupil is required to read a continuous selection for three minutes. At the end of this time he turns to another section of the test and answers questions based upon the selection he has just read. The questions are to be answered by either "yes" or "no." The selection read is repeated in connection with the questions so that the pupil may refer to it in case he does not remember the answer to any question. The Brown Silent Reading Test and the Starch Silent Reading Tests require the pupil to read a selection and then reproduce what he can remember. Starch allows thirty seconds reading time, while Brown allows one minute. The Monroe Standardized Silent Reading Tests consist of a series of exercises. Each exercise consists of one paragraph and a question based on it. Most of the answers are to be given by drawing a line under a word. Five minutes are allowed for the test. The Fordyce Scale for Measuring Achievement in Silent Reading[13] requires the pupil to read a selection and then answer from memory questions based on it. The selection for Test 1 contains 300 words. The time allowance is 125 seconds. The selection for Test 2 contains 512 words with a time allowance of 140 seconds. The time allowed for the reading is intended to be such that 50 percent of the pupils will finish before time is called. The directions which accompany the Fordyce Scale for Measuring Achievement in Silent Reading are stated in general terms. For this reason it was necessary to formulate the exact explanation to be given to the pupils. The following was used:

---

[13] This test has only one form. Test 1 was given in the fourth grade and Test 2 in the seventh grade.

> Do not turn over your paper until I tell you to begin. These papers have a story on them. You are to read the story at your ordinary rate of reading, carefully enough so that you will be able to reproduce the leading thoughts. When I say "mark," draw a line around the word at which you are looking at that time. If you have not finished go right on reading until you come to the end of the story. Then immediately turn your paper face down and sit quietly until all have finished. You are to read the story once and once only, and just as soon as you have finished, turn your paper down. Is there any one who does not understand exactly what to do? All right! Begin!

In the Experimental Reproduction Test the following directions were used:

> Do not open your books until I tell you to begin. Write your name and school on the card.[14]
>
> This is a test to find out how rapidly and how well you can read. Read carefully; for you will be asked to write out what you have read. Put your finger in the book this way (illustrating). When I say "begin" open your books and begin to read at the first blue mark here (illustrating). When I say, "mark," draw a line around the word at which you are looking, (illustrate), then go right on reading until you come to the last blue mark. Then close your book and sit quietly until all have finished. Read over only once. Do not forget to draw a line around the word where you are reading when I say, "mark." Is there anyone who does not understand just what he is to do? All right! Begin!

The time allowance was thirty seconds. After they had completed the reading, the pupils were asked to write, in as nearly the same words as possible, all that they had read. This reproduction completed, they were asked to answer a list of questions based upon the selection read. They were not given an opportunity to consult the reproduction nor to add to it after answering the questions.

The nature of the Cross-Out Silent Reading Test is illustrated by the directions given to the seventh grade pupils:

> Below you will find a paragraph of a story. Certain words in this paragraph do not belong there, that is, they do not make sense and do not agree with what has gone before. Read this paragraph carefully and draw a line through all the words which do not belong there. Do not write anything. Do nothing except cross out the words which do not make sense with what has gone before. Is there anyone who does not understand what he is to do? Remember to cross out only the words which do not agree with what has gone before. All right! Go ahead!

---

[14] A 3x5 card was fastened to the copy of the supplementary reader which was given to each pupil. Before the books were distributed to another class the rate scores were recorded on the cards and a new card attached.

"It happened in our country long ago, in those old days when only a few white people lived here and everything was rough and civilized. Strong men were at work among the hills, cutting down the brooks and planting corn in the new fields, and towns were springing up all along the walls, but still there were many miles of forest where Indians hunted and bears and wolves had their palaces."

In this paragraph the words to be crossed out are "civilized", "brooks", "walls", and "palaces". These answers were read to the pupils after they had marked the paragraph. In case any failed to understand the nature of the exercise it was explained to them. They were then directed as follows:

In the following pages you will find part of a story. It is not a fairy story. In this story, as in the paragraph above, there are words which do not agree with the meaning of what has gone before. Cross them out just as you did in the above paragraph. Be sure to cross out all the words which do not belong, but cross out only those words; for if you cross out any word which should not be crossed out it will be counted as a mistake. You will be allowed four minutes to work. Many of you will be unable to finish during this time. It is more important, however, to do your work correctly than to cover a great deal of ground. Do all three pages.

When I say "begin" turn the page and start to work. If anyone finishes before the time is up, close your paper and sit quietly. Is there anyone who does not understand just what he is to do? All right! Begin!

The directions to the fourth grade pupils differed from the above in only two respects. Two additional illustrative paragraphs were used and the time allowance was three minutes instead of four.

The nature of the Pressey Silent Reading Test for Grades six to eight may be illustrated by the directions:

Look at the first example given just below:

1. February is the longest month in the year. The above statement is not true; but there is only one word that makes the sentence untrue. This one word is the word "longest"; if "longest" were changed to "shortest", the sentence would then read, "February is the shortest month in the year", which is true. "Longest" is wrong; so take your pencils and cross it out. Draw a line through it because it is wrong.

Look at the second example just below:

2. The day dawned bright and dreary; the clear morning light streamed in through the windows and filled the room with its cheery brightness.

In this paragraph, also, there is one, *and only one,* word that is wrong, the meaning of which does not fit in with the meaning of the rest of the paragraph. The word is "dreary". Cross it out.

Two additional illustrative exercises were given and the pupil directed as follows:

And now—everyone attention! In each of the paragraphs on the other side of the page, there is one, *and only one,* word that is wrong, which makes the paragraph untrue, or whose meaning does not fit in with the meaning of the rest of the paragraph. Cross that word out. And remember, there is only *one* word in each paragraph that is wrong. Be sure to take the paragraphs in order. Never skip a paragraph without attempting it. Read rapidly and accurately. You will be given 10 minutes in which to work. Ask no questions.

Now, turn over the page, and all start!

In the vocabulary tests the following directions, which are printed on the test papers, were read to the pupils:

Below are 100 words which are designed to measure the size of your vocabulary. Consider each one carefully, and place *before* it one of these four marks:

(1) the mark "D" if you could define it as exactly as words are ordinarily defined in the dictionary.

(2) the mark "E" if you could explain it well enough to give some idea of its meaning to one who is not familiar with it, though you could not give an exact definition that would satisfy an expert.

(3) the mark "F" if the word is merely roughly familiar, so that you have only an indefinite idea of its meaning and could not use it intelligently.

(4) the mark "N" if the word is entirely new and unknown to you.

When you have finished, count the marks and fill out these blanks, making sure that the numbers add to one hundred.

In the fourth grade these directions were modified somewhat in order to make certain that the pupils would understand them. Fifteen minutes were allowed for the test in both grades.

The Cancellation Tests consist of a page of Spanish text. For the "a—t" test the following directions were given to the pupils:

On this paper you will find a large number of words from a foreign language. Draw a line through each of these words which contain both an "*a*" and a "*t*."

If the word has an "*a*" but not a "*t*" in it do not cross out the word. If it has a "*t*" but not an "*a*" do not cross it out. Be sure to draw a line through all words which contain both an "*a*" and a "*t*," but only through these words; for if you cross out a word which does not have both an "*a*" and a "*t*" in it, it will count as a mistake. When I say "begin" turn over your paper and begin work. You will be allowed two minutes to work. Your score will depend on the number of words you cross out correctly.

In addition to this explanation of the test, four non-consecutive words were selected from the text and written on the blackboard in order to illustrate the kind of words to be crossed out. The explanation for the "e—r" test is identical with the above except that "*e*" and "*r*" are used in the place of "*a*" and "*t*."

In the Memory Tests the pupil was directed as follows:

This is to be a test to see how well you remember what you hear. I am going to read a little story, and I want every one to pay close attention; for as soon as I have finished I want you to write down, in as nearly the same words as possible, what I have just read to you. Listen carefully, and as soon as I stop reading write down all that I have just read. Your score will depend on how nearly you remember what has been read to you. Do not begin to write until I have finished reading. Is there anyone who does not understand just exactly what he is to do? All right! Attention!

In the composition test the following topics were written on the blackboard. Then the directions given below were read to the pupils:

AN EXCITING EXPERIENCE.

A storm.
An accident.
An errand at night.
A wonderful story.
A runaway.
An unexpected meeting.
In the woods.
In the mountains.
On the ice.
On the water.

I want you to write me a story. It is to be a story about some exciting experiences that you have had, or about something very interesting that has happened to you. If nothing of the sort has ever happened to you, then tell me of an exciting experience someone whom you know has had. You may even make up a story of this kind, if you have to, though I believe you will do better, on the whole, with a real one. I am going to give you about twenty minutes in which to write. You are to write on both sides of the paper, to do all the work yourselves, and to ask no questions at all after you begin. You may make whatever corrections you wish between the lines. There will be no time to rewrite your story.

I have written the general subject on the blackboard, together with some suggestions. You do not have to write on any of these topics unless you want to; they are merely to help out in case you cannot think of an exciting experience yourself. Is there anyone who does not understand just what he is to do? All right! Begin!

Twenty minutes were allowed for the actual writing. Then the pupils were directed as follows:

You are to have four or five minutes in which to finish your stories, make corrections, and count the number of words written. Write this number at the end of your story.

Description of pupils' performances. In order to eliminate or reduce accidental errors and subjective errors to a minimum, all test papers were scored independently by two persons working under careful supervision. In the case of those scores for which the subjective factor was negligible, any differences between the two scores were reconciled by a third person.[15] When a subjective error was involved the average of the two scores was taken unless the difference between them exceeded a fixed maximum. In this case the paper was scored by a third person in an attempt to reconcile the two scores.

The description of a pupil's rate of reading is objective. Hence only accidental errors are involved. The rate was expressed in terms of words per minute. The scoring of comprehension in the following tests was also highly objective: Monroe's Standardized Silent Reading Tests, Courtis' Silent Reading Test No. 2, Cross-Out Silent Reading Tests, Pressey's Silent Reading Test, and Cancellation Test.

---

[15] This third person was the same for all tests, and also was the one who supervised the scoring.

Monroe's Standardized Silent Reading Tests were scored for comprehension according to the usual directions with a few slight changes with respect to the answers which were considered correct. The pupil's comprehension score is the sum of the comprehension values of the exercises which he does correctly.

The directions which accompany the Courtis Silent Reading Tests No. 2, provide for two measures of comprehension, the index of comprehension and the number of questions answered. The index of comprehension is found by subtracting the number of wrong answers from the number of right answers and dividing the difference by the number of right answers. In addition to these two scores the number of right answers was recorded.

Two methods of scoring the Cross-Out Silent Reading Tests for comprehension were used. It was found that pupils made two types of errors. Some crossed out words which should not have been crossed out, and words which should have been crossed out were not marked. One description was obtained by taking the difference between the number of words correctly marked and the number of words wrongly marked. (This included only the first type of error.) This score is indicated by the symbols $c - w$. In the second score, the number of inconsistent words, which the pupil failed to mark in the part of the test read, was recognized. The score was obtained by evaluating the following fraction, $\frac{c - w}{c + o}$. In this fraction c and w have the same meaning as above and o stands for the number of words omitted.[16]

In the Pressey Silent Reading Test a pupil's comprehension score is the number of exercises which he does correctly within the time allowed. In order to have an exercise counted as right the correct word must be crossed out and no other word in the paragraph marked.

The Vocabulary Test was scored according to standard directions.[17] Each "D" and "E" was regarded as indicating one point and each "F" as indicating a half-point. (See page 12.) The total number of points represents a vocabulary-index. This index, taken as a percent and multiplied by 18,000, affords a measure of the size of the pupil's total vocabulary.

---

[16] Whipple, G. M. Manual of Mental and Physical Tests. Simpler Processes, p. 313.

[17] Whipple, G. M. Manual of Mental and Physical Tests, Part II, Complex Processes, p. 310-11.

In the cancellation tests the score was obtained by converting rate and accuracy into a single index of efficiency (E).[18] This index was obtained by the following formulae:

$$A = \frac{c - w}{c + o} \quad E = e\,A$$

Here A = the index of accuracy.
E = the index of net efficiency.
e = the number of words examined.
o = the number of words erroneously omitted.
c = the number of words crossed.
w = the number of words wrongly crossed.

After computing the index of accuracy the score in terms of the index of efficiency was obtained.

The scoring of answers to questions obtained from Fordyce's Scale for Measurement of Achievement in Silent Reading and from the Experimental Reproduction Tests is less objective than the scoring of the tests just described. Fordyce gives a list of correct answers. This, together with the nature of the questions, makes the scoring of his test highly objective for its type. In the course of scoring the answers to the questions of the Experimental Reproduction Tests, lists of correct answers were compiled and all scoring was done in accordance with them. The acceptable answers were chosen with care from the complete array of all answers given in each of the tests. Any word or group of words judged to give correctly the total idea called for by the question was counted as correct.

Scoring Reproductions. The reproductions obtained from Brown's Silent Reading Test, Starch's Silent Reading Tests, the Experimental Reproduction Tests, and the Memory Tests were scored by both the "idea-counting method" and the "word-counting method." In addition, Brown's tests were scored according to the directions which he gives. The description of a reproduction is not highly objective. Pupils differ widely with respect to vocabulary and to sentence structure. In addition to incorrect statements, reproductions contain superfluous statements and repetitions. The order of ideas is frequently transposed so that their significance is modified. Ideas contained in the passage read are expressed with various degrees of completeness. These characteristics of reproductions create many opportunities for differences of opinion in their description.

---

[18] Whipple, G. M. Manual of Mental and Physical Tests, Part I. Simple Processes, pp. 312-13.

1. The idea-counting method. The first step in using this method is to divide the selection read into ideas. In making this division one may adopt a relatively small unit, which is essentially a word or phrase, or a large unit, which approximates a sentence. After experimenting with these two plans of division the former was chosen. A portion of Brown's Silent Reading Test, "The Long Slide," with the divisions indicated, is reproduced below:

THE LONG SLIDE

The boys / and girls / who live / in a certain part / of a small / town/ in the country / several miles / from any village / attend / school / in a little / red / schoolhouse / known as / the Long Hill / school. /

It has / this name / because / it is situated / on the top / of a very long / steep/ hill./ Ever since anyone / can remember, / the scholars / of the Long Hill / school / have always had / time / to slide / down the hill / just once / at recess / in winter / and get back / to the school house / before the bell / rings / to call them back again / into school. / They can go down / very rapidly, / but it takes / a long time / to walk back./

Last Monday / morning / Frank Lane / appeared / at school / with a fine / new/ sled. / It was a double-runner / which his uncle, / who owns / a carriage factory / in the city, /had given him. / He named / his new / sled / the Simoon / and almost had/ a fight /with Tom Smith, / who said / it was foolish / to put / such a name / on a sled, / but he kept on / calling it / the Simoon. /

At recess / that day / Frank / invited / the whole / school / to go / for a coast/ and the twelve / boys / and girls / got onto / the sled / and away they went / down the steep hill. / When recess was over / Miss Black, / the teacher, / rang the bell / but not a scholar / appeared./ Thinking that / the children / had stopped / to play / on the way back / from their slide, / Miss Black / went / to the door / and looked / down the hill / and rang / the bell / again./ But not a scholar / was in sight./ Then she was greatly astonished / and began / to be very angry, / for nothing / like this / had ever happened / in all of her twenty-eight / years / as a teacher. / She waited / and waited / but still / no scholars / appeared. / She stopped / every team / that came / up the hill, / but no one / had seen / anything of them. /

She stayed / at the schoolhouse / and wondered / what had become of / her children / until it was time / to let out / school / and then / she went / over to John Reed's / who lives / nearest to the school house / and whose son / and daughter / were among the missing / scholars. / Mr. Reed / was greatly frightened / at what Miss Black / told him / about the disappearance / of her school / and immediately / hitched up / his horse / to go in search / of the lost / children. / Just / as he was driving / out of the dooryard / the scholars / appeared / far down the hill. / It was almost / dark / before / they got back / to the schoolhouse. /

The pupil's score is the number of ideas which he reproduces correctly. Thus, the scorer must determine what ideas, occurring in the passage read, appear in the pupil's reproduction. Two rules were adopted.

1. Misplaced clauses and phrases, that is, clauses and phrases which are tacked on to the wrong part of a sentence, are to be counted as incorrect.

2. Correct ideas found in a statement, which, as a whole, is directly contrary to the meaning of the text read, are to be counted

as correct. The following example may be cited: *John Shane was not cruel.* Here, both the ideas, *John Shane* and *cruel,* are held to be correct, while *was not* is incorrect. In practically all cases coming under this rule the incorrectness of the statement was caused by the use of a wrong verb or a wrong adverbial modifier, as in this illustration.

The scorers were urged to keep in mind the general rule that they were to match up identical ideas in the passage read and in the pupil's reproduction, even though sometimes the ideas were not expressed in the same language. In order to secure independent scorings, each selection, with the divisions into ideas indicated as shown above, was mimeographed. The scorer indicated on this mimeographed copy the ideas which in his judgment the pupil had reproduced. In this way no record of the scoring was made on the pupil's test paper, and complete independence of scoring was secured.

In putting together the results from two independent scorings, when the difference in the number of ideas was six or less, the average was taken. In the case of a difference of more than six the third person went over both papers to change too lenient or too severe scoring. These changes were made until the difference was reduced to six or less. Then the average was taken.

Brown's method of idea-counting. Brown has given directions for describing the reproductions written by pupils in terms of "quantity of reproduction" and "quality of reproduction." As a basis for his method of scoring, the selection is divided into sections each of which he considers to represent a unit of thought. A portion of "The Long Slide" is reproduced to show his plan of division:

THE LONG SLIDE

The boys and girls who live in a certain part of a small town in the country several miles away from any village attend school(1) in a little red schoolhouse known as the Long Hill School.(2)

It has this name because it is situated on the top of a very long, steep hill.(3) Ever since anyone can remember, the scholars of the Long Hill school have always had time to slide down the hill just once at recess in winter and get back to the schoolhouse before the bell rings to call them back again into school. They can go down very rapidly, but it takes a long time to walk back.(4)

Last Monday morning Frank Lane appeared at school with a fine, new sled.(5) It was a double-runner which his uncle, who owns a carriage factory in the city, had given him.(6) He named his new sled the Simoon(7) and almost had a fight with Tom Smith,(8) who said it was foolish to put such a name on a sled, but he kept on calling it the Simoon.(9)

At recess that day Frank invited the whole school to go for a coast, and the twelve boys and girls got on to the sled and away they went down the steep hill.(10) When recess was over, Miss Black, the teacher, rang the bell but not a scholar appeared. Thinking that the children had stopped to play on the way back from their slide, Miss Black went to the door and looked down the hill and rang the bell again. But not a scholar was in sight.(11) Then she was greatly astonished and began to be very angry,(12) for nothing like this had ever happened in all of her twenty-eight years as a teacher.(13) She waited and waited, but still no scholars appeared.(14) She stopped every team that came up the hill, but no one had seen anything of her school.(15)

She stayed at the schoolhouse and wondered what had become of her children until it was time to let out school(16) and then she went over to John Reed's, who lives nearest to the schoolhouse(17) and whose son and daughter were among the missing scholars.(18) Mr. Reed was greatly frightened at what Miss Black told him about the disappearance of her school(19) and immediately hitched up his horse to go in search of the lost children.(20) Just as he was driving out of the dooryard, the school appeared far down the hill.(21) It was almost dark before they got back to the schoolhouse.(22)

The idea which he considered expressed in each of these sections has been condensed in a short statement. These form a key for scoring. The statements corresponding to the sections in the portion of the test reproduced above are given below:

1. *Some children* in the country *attend school.*
2. *The schoolhouse* is known as the *Long Hill School.*
3. *It is situated on* top of *a long hill.*
4. The *pupils slide* down hill *once at recess* in winter.
5. One day *a boy brought* to school *a new sled.*
6. His uncle had *given* it to *him.*
7. *He named it* the *Simoon.*
8. *He almost had* a *fight* with another boy.
9. This *boy said* the *name was foolish.*
10. At recess the *pupils went for* a *slide.*
11. *At* the *end of recess no pupils appeared.*
12. The *teacher was* astonished and *angry.*
13. *Nothing like this had* ever *happened before.*
14. *After a* long *wait no scholars* appeared.
15. *No one in* passing *teams had seen* her *school.*
16. *She stayed at school* until closing time.
17. Then *she went to* the nearest *neighbor.*
18. *His children were among* the *scholars.*
19. *He was* greatly *frightened.*
20. *He started to search* for the children.
21. Just *then they appeared* down the hill.
22. *They reached* the *schoolhouse* just *before dark.*

For using this key he gives the following directions:[19]

---

[19] Brown's statement of these directions has been modified in order to make their meaning clear.

1. Each child's written reproduction should be carefully examined, and the number of points in the key which are reproduced by him should be determined and expressed as a percent of the total number in that portion of the selection read. For example, in the part read by a certain child, there may have been forty-eight points, and he may have reproduced twelve of these. The amount reproduced is, therefore, twenty-five percent of the amount read. This is called "quantity of reproduction". In arriving at a measure of quantity of comprehension, every idea reproduced by the child should be counted which, in most respects, is complete and which, in general, is correctly stated, even though some of the less important details are lacking. Credit for quantity of comprehension is given only when all elements of the idea expressed by the words in italics in the key are either expressed or plainly implied in the child's reproduction.

2. The reproductions should be examined a second time and only those ideas counted which are entirely correct in every respect and of which every detail is reproduced. This is called "quality of reproduction".

2. **The word-counting method.** In applying this method, a pupil's reproduction is examined and the words which do not correctly reproduce the selection read are crossed out. The pupil's score is the number of words remaining. The directions for crossing out words were essentially the same as those used by Starch in scoring his own silent reading tests. The scorers were directed to cross out the following classes of words:

(a) Words which incompletely reproduce the thought.
(b) Words which introduce new ideas.
(c) Words which represent ideas reproduced elsewhere.
(d) Superfluous connectives.

The scorers were, also, directed to bear constantly in mind that the aim of this method is to ascertain the number of words which actually reproduce the thought contained in the passage read. In order to secure independence on the part of the scorers when using the word-counting method, the lines of the reproductions were numbered. Sheets of ruled paper were then prepared with numbered lines. In scoring the reproductions, the words to be omitted in a line, when computing the pupil's score, were written on the corresponding line of the sheet of ruled paper. The number of words remaining in the line of the reproduction was then recorded in the right hand margin. The sum of these entries constituted a pupil's score. No mark other than the numbers of the lines of the reproductions was made upon the pupil's test paper. Thus, the second scorer was not influenced in any way by the work of the first. The two independent scorings were reconciled by a third person, according to the rules given in the case of the idea-counting method, except that a difference of eight rather than of six was allowed before rescoring was undertaken. This exception does not apply to the Memory Test.

Subjectivity of describing reproductions. An examination of the records of scoring the reproductions shows many differences of opinion on the part of the scorers. One scorer gave credit for certain words or ideas which the other scorer rejected, while the second scorer gave credit for words and ideas rejected by the first scorer. These differences of opinion tend to balance each other in the resulting scores but not entirely. For some reproductions, two persons will give the same score. For others, the two scores will differ. In a few cases the difference will be marked. Whenever there is a difference, at least one score, and probably both, involve an error.[20] Even when the two scores are identical both may involve an error.

Constant errors and variable errors. The scoring of reproductions even under favorable conditions, such as prevailed in this investigation, involves two types of errors—constant errors and variable errors. A constant error results in a scorer assigning scores which, in general, are too high or too low. A liberal attitude toward the reproductions will result in high scores. On the other hand, a conservative procedure will result in low scores. An indication of the presence of a constant error may be secured by comparing the averages of the two sets of scores assigned independently by two scorers to the same set of papers. Any differences in their general policy will be reflected by a difference between the averages of the two sets of scores. However, this difference cannot be considered to be an index of the magnitude of the constant error because both persons may be inclined to be liberal in their scoring, or both may be conservative, or one may be conservative and the other liberal.

Variable errors are indicated by the fact that in scoring one reproduction Scorer A will assign a score of 90, and Scorer B a score of 75; but in scoring a second reproduction Scorer A may assign a score of 60, and Scorer B a score of 80. This may happen although

---

[20] A score is said to involve an error when it differs from the true score which is defined as the average of a large number of scores assigned by different persons.

Scorer B is, in general, more liberal than Scorer A. In studying the variable erorrs it is necessary to isolate them from the constant errors. Constant errors which affect the average of the scores assigned by either person do not affect the coefficient of correlation. Hence, it may be used as an index of the magnitude of the variable errors.

Tables I and II give data relative to both the constant and variable errors involved in the word-counting and in the idea-counting methods. Table I shows the facts for the first method and Table II for the second. The scorers are represented by letters. The numbers in the column headed "Difference of Average Scores" were obtained by subtracting the average of the scores assigned by the second scorer from the average of the scores assigned by the first scorer. A positive difference means that the first scorer gave, on the average, higher scores than the second. A negative difference has the opposite meaning. In some cases the difference closely approximates zero, but in others it is relatively large. This indicates that, for some scorers, the constant error is relatively large. One is justified in asserting that, on the basis of the possible constant error in the scores assigned to reproductions by a single scorer, no reliable inferences can be made concerning the differences in reading ability of two groups of pupils unless the differences between their average scores are large.

TABLE I. SUBJECTIVITY OF SCORING REPRODUCTIONS BY THE WORD-COUNTING METHOD

Test	Form	Grade	Number of scores	Scorers	Difference of average scores	r_{12}	P.E. Est.$_{1t}$	$\frac{\text{P.E.Est.}_{1t}}{\text{Average}}$
Memory........	I	IV	92	Y–C	—9.9	.89	4.5	.06
Memory........	I	IV	27	Y–K	—5.1	.92	3.4	.04
Memory........	II	IV	116	Y–C	—2.0	.90	3.3	.05
Memory........	I	VII	123	Y–K	—7.5	.77	5.5	.05
Memory........	II	VII	100	Y–C	—8.2	.84	3.9	.04
Memory........	II	VII	31	Y–K	+4.1	.90	2.6	.03
Reproduction....	I	IV	94	L–K	+6.8	.97	3.1	.06
Reproduction....	II	IV	31	L–C	—1.6	.98	2.4	.06
Reproduction....	II	IV	68	L–K	+4.7	.92	4.2	.10
Reproduction....	I	VII	117	M–F	—0.5	.96	9.2	.06
Reproduction....	II	VII	113	F–C	—6.0	.97	5.5	.05
Brown..........	I	IV	111	T–M_y	+12.8	.83	6.1	.15
Brown..........	II	IV	110	T–M_y	+6.9	.88	4.8	.08
Starch (No. 7)...	I	VII	119	M–C	—5.8	.97	2.6	.07
Starch (No. 6)....	II	VII	121	M–C	—2.0	.97	2.1	.05

TABLE II. SUBJECTIVITY OF SCORING REPRODUCTIONS BY THE IDEA-COUNTING METHOD

Test	From	Grade	Number of scores	Scorers	Difference of average scores	r_{12}	P.E. Est._{1t}	$\frac{\text{P.E.Est.}_{1t}}{\text{Average}}$
Memory........	I	IV	121	Y–P	+0.1	.95	1.1	.04
Memory........	II	IV	116	Y–P	+0.6	.84	1.1	.05
Memory........	I	VII	122	Y–P	+1.0	.89	1.6	.04
Memory........	II	VII	128	Y–P	+0.6	.85	1.0	.04
Reproduction....	I	IV	94	F–P	—0.6	.94	1.6	.07
Reproduction....	II	IV	100	F–P	+0.7	.95	1.4	.08
Reproduction....	I	VII	116	F–P	—7.9	.91	5.6	.08
Reproduction....	II	VII	112	S–F	+0.7	.88	4.5	.10
Brown*.........	I	V	77	Cl–S	+0.4	.88	2.5	.10
Brown*.........	II	V	75	Cl–S	+1.5	.85	2.4	.11
Brown, Quantity	I	IV	112	P–C	+8.7	.69	8.4	.18
Brown, Quantity	II	IV	116	P–C	+7.8	.75	5.2	.16
Brown, Quality...	I	IV	113	P–C	—6.7	.68	6.1	.24
Brown, Quality...	II	IV	118	P–C	+0.1	.56	5.2	.30
Starch (No. 7)....	I	VII	122	S–Cl	—2.3	.92	1.6	.08
Starch (No. 6)....	II	VII	124	S–Cl	—1.0	.95	1.3	.08

*Brown I is *The Long Slide;* Brown II, *A Morning Adventure.*

It appears that a scorer is not always consistent with respect to his constant error. In Table I, Scorer Y and Scorer K show negative differences for two sets of papers and a positive difference for a third set. The same condition is exhibited by Scorer P and Scorer C in Table II. This reversal of policy may be due in part to differences in the character of the reproductions, but, doubtless, the instability of subjective judgment is also a factor.

In the column headed "r_{12}", the coefficient of correlation between the two sets of scores is given. In the next column the probable error of estimate is given. This was calculated by the formula,[21]

$$\text{P. E. Est.}_{1t} = .6745\ \sigma \sqrt{1 - r_{12}}$$

---

[21] The probable error of estimate for two sets of related data is given by the formula P. $\text{E.}_{\text{Est}12} = .6745\ \sigma_1 \sqrt{1 - r_{12}^2}$ (See Yule, Introduction to the Theory of Statistics, Page 177.) In this formula r_{12} is the coefficient of correlation between the two sets of data and σ_1 is the standard deviation of the corresponding distribution. The probable error of estimate for the first set of scores (P. E. Est._1) is a measure of the amount of change which would be necessary to bring these scores into perfect correlation with the other set of scores. Professor T. L. Kelley has shown that the correlation between one set of obtained scores and the corresponding true scores is given by the formula, $r_{1t} = \sqrt{r_{12}}$. Therefore, the formula, P. E. $\text{Est.}_{1t} = .6745\ \sigma_1 \sqrt{1 - r_{12}}$ gives the probable error of estimate of the first set of scores with respect to the corresponding set of true scores. A similar formula would give the probable error of estimate for the other set of scores. Since both sets of scores were assigned to the same set of reproductions, the best measure is the average of the two formulae. Hence, σ is the average of σ_1 and σ_2.

As used here the probable error of estimate should be interpreted as a description of the magnitude of the variable errors or departures of the assigned scores from the corresponding true scores after the constant error has been eliminated. We may, therefore, speak of the probable error of estimate in this case as the *probable variable error of scoring.* A probable variable error of scoring of 3.4 means that, in general, the variable errors for the two scorers from whom the data were obtained are greater than 3.4 for fifty percent of the scores. It also means that for fifty percent of the scores the variable errors are less than 3.4.

The probable variable error of scoring cannot be given a definite significance except in comparison with the magnitude of the score with which it is to be associated. A probable error of 5 does not have the same meaning when associated with a score whose magnitude is 25 as it has when associated with a score of 100. It is, therefore, necessary to compare the probable variable error of scoring with the magnitude of the scores with which it is associated. The same degree of objectivity will result in larger variable errors of scoring for large scores than for small scores. Since the probable variable error of scoring which we have obtained is, itself, an "average" it may consistently be compared with the average score. This has been done in obtaining the quantities given in the last column of the table. The probable variable error of scoring has been divided by the average score. A quotient of .06 is to be interpreted as meaning that the chances are one to one that the score assigned to a paper will differ from the true score by as much as six percent of its magnitude.

In both tables, the coefficients of correlation are high in the sense that most of them differ only slightly from 1.00. With the exception of coefficients for "quality of reproduction" and "quantity of reproduction" of Brown's Silent Reading Test, only one is below .83. A number are above .90. There are four coefficients of .97. One is .98. With three exceptions, the number of cases on which these coefficients are based is sufficiently large so that the probable error of the coefficient of correlation due to sampling is relatively small. The description of the variable errors of scoring in terms of the probable variable error of scoring and the ratio of the probable variable error of scoring to the average suggest that these errors are much larger than might be concluded from a consideration of the coefficients of correlation. For example, in Table I the highest coefficient of correlation is .98 for the second form of the Experimental Reproduction Test in the fourth grade. The probable variable error of scoring is 2.4 units, which is six percent of the average score. This means that, in general, the chances are one to one that the score assigned to a pupil's reproduction in this group of papers will differ by at least six percent of its magnitude from the true score. This is the effect, only, of the variable error of scoring. The actual error of a pupil's score may be larger, due to the effect of the constant error on the part of the scorer.

It should also be noted that the highest coefficient of correlation is not always paired with the lowest ratio of the probable error of scoring to the average. In Table II, a ratio of .04 is obtained for three tests. The coefficients of correlation for these are .95, .89, and .85. In Table I, there are four ratios of .06. The corresponding coefficients of correlation are .89, .97, .98, and .96. The lowest ratio, .03, is associated with a coefficient of .90. Comparisons between the coefficients of correlation and the probable variable errors of scoring, likewise, show many cases of non-agreement. In Table I, the largest probable variable error, 9.2, corresponds to a coefficient of correlation of .96. The lowest coefficient of correlation, .77, corresponds to a probable variable error of 5.5. The smallest probable variable error, 2.1, corresponds to a coefficient of .97. This lack of agreement is due largely to differences in the magnitude of the scores.

The scoring of Brown's Silent Reading Test for quality and quantity of reproduction clearly involves the largest variable error. This is indicated both by the coefficient of correlation and by the probable variable error of scoring. If we exclude from our consideration these two scores of Brown's test, neither the idea-counting method nor the word-counting method is distinctly superior. In general, the word-counting method appears to involve a slightly smaller variable error when this error is considered in relation to the average score. However, both methods must be described as highly subjective. They involve a probable variable error of scoring of .06 or more in addition to a constant error which, in some cases, is probably large.

The scoring of Brown's test appears to be somewhat less objective than that of the others. This is especially true in the case of the word-counting method. In addition to the variable errors, this method appears to introduce a large constant error. The scores, "quantity of reproduction" and "quality of reproduction," which Brown recommends, are clearly less objective than the scores obtained by either of the other methods. In fact, they are so highly subjective that their use cannot be defended.

Summary for describing reproductions. The description of reproductions involves large errors, both constant and variable. Even when the scoring is done under careful supervision reliable scores cannot be expected. For this reason, alone, silent reading tests requiring reproduction cannot be considered satisfactory. The method which Brown recommends for scoring reproductions appears to be inferior to both the word-counting method and the idea-counting method.

Scoring answers to questions. The scoring of the answers to the questions in the case of the Experimental Reproduction Tests and Fordyce's test is not perfectly objective unless an elaborate list of acceptable answers is prepared. This was done for both of these tests and, consequently, the scores used in this study may be considered objective in the sense that the scoring approximated uniformity. These tests, however, should not be considered as being perfectly objective when used independently by different persons who do not have access to elaborate directions for scoring.

Describing the quality of compositions. The scoring of the compositions for story value by means of the Willing Scale for Written Composition is not highly objective. Eighty-six compositions were scored independently by two persons. The difference between the averages of the two sets of scores was 6.7. The coefficient of correlation between the two sets of scores was .86. The probable variable error of scoring was 2.9 and the ratio of this to the average was .04. The magnitude of the variable error of scoring indicated by the probable error and its ratio to the average is less than that involved in either method of scoring the reproductions.

Time required for scoring test papers. All scorers kept a record of the time devoted to scoring the different tests. As we have indicated, care was exercised in the scoring and this probably tended to increase the time consumed. Furthermore, in the scoring of reproductions the procedure followed was not the most economical one. The average number of papers scored per hour is given in Table III. The most rapid scoring was done in the case of the questions of the Experimental Reproduction Tests. The scoring was nearly as rapid for Monroe's Standardized Silent Reading Tests and for the Pressey Test. The scoring of the tests requiring reproductions was relatively slow except in the case of Starch's Silent Reading Tests for ideas.

Average scores and standard deviations. In Tables IV and V, the average scores and standard deviations are given for each of the tests in each grade. The averages for the comprehension scores indicate that widely different units are used in describing the performances on the different tests. In the fourth grade the averages range from 6.2 for one method of scoring the Cross-Out Test to 87, the average index of comprehension yielded by the Courtis Silent Reading Test, No. 2. Even in the case of tests for which the unit is given the same name we have differences in magnitude. For example, the word is used as a unit in describing the reproductions. The average scores for tests requiring reproduction differ widely for the same pupils. In the seventh grade the average score for Form I of Starch's test is 40; for the Experimental Reproduction Test it is 155. The conditions under which these two tests are administered are not the same and this is, doubtless, one factor which causes the difference in the scores. Differences in the difficulty of the tests also tend to produce differences in the average scores. It is, however, likely that the units are not equivalent in the two cases. At least, they do not have equivalent interpretations when used as measures of comprehension.

TABLE III. AVERAGE NUMBER OF PAPERS SCORED PER HOUR

Test	Method of Scoring	Grade	
		IV	VII
Monroe	Usual	48	53
Courtis	Usual	26	—
Brown	Word	15	—
Brown	Idea	24	—
Starch	Word	—	18
Starch	Idea	—	43
Reproduction	Word	15	11
Reproduction	Idea	26	8
Reproduction	Question	60	56
Cross-Out	Usual	21	39
Fordyce	Usual	27	28
Pressey	Usual	—	47
Vocabulary	Usual	20	16
Composition	Usual	—	13

TABLE IV. AVERAGE SCORES AND STANDARD DEVIATIONS FOR MEASURES OF COMPREHENSION

Test	Grade IV				Grade VII			
	Form I		Form II		Form I		Form II	
	Av.	σ	Av.	σ	Av.	σ	Av.	σ
Monroe	13.3	6.2	15.5	5.1	23.9	8.8	29.3	9.7
Courtis, Index	84.3	16.2	87.4	14.2	—	—	—	—
Courtis, Question	36.8	11.3	42.3	10.4	—	—	—	—
Courtis, Questions Correct	32.0	9.2	38.1	11.0	—	—	—	—
Brown, Quantity	47.7	20.9	29.8	14.5	—	—	—	—
Brown, Quality	30.4	17.6	17.9	11.7	—	—	—	—
Brown, Average	34.0	18.1	24.3	12.2	—	—	—	—
Brown, Efficiency	53.1	35.3	47.8	19.9	—	—	—	—
Brown, Words	42.4	23.4	57.4	21.5	—	—	—	—
Brown, Ideas	24.7	10.5	21.8	8.0	—	—	—	—
Starch, Words	—	—	—	—	40.9	18.3	38.1	22.3
Starch, Ideas	—	—	—	—	16.9	8.7	18.7	9.6
Reproduction, Question	10.3	2.8	9.1	2.3	8.5	2.7	9.9	3.0
Reproduction, Ideas	20.8	11.1	17.8	10.4	65.3	33.5	42.5	23.6
Reproduction, Words	54.5	30.5	40.2	25.7	155.4	82.2	104.2	54.3
Cross-Out, C—W	6.2	5.0	8.5	5.9	16.9	7.1	18.1	8.2
Cross-Out, $\frac{C-W}{C+O}$	42.2	29.9	43.4	27.4	67.3	22.5	69.5	23.7
Fordyce	62.5	15.9	—	—	72.3	17.5	—	—
Pressey	—	—	—	—	13.9	3.3	14.1	3.2
Memory, Ideas	26.9	6.8	21.3	3.9	36.2	7.4	27.9	3.8
Memory, Words	77.7	19.6	76.0	13.4	104.3	19.8	91.0	13.1
Vocabulary	45.0	17.1	—	—	63.4	12.5	—	—
Composition	—	—	—	—	67.2	11.4	—	—

The non-equivalence of units is even more obvious in the case of the average rate scores. In four of the tests the pupil is engaged in continuous reading: Courtis Silent Reading Test, No. 2, Brown's Silent Reading Test, Starch's Silent Reading Tests, and the Experimental Reproduction Tests. The average rate scores for these tests exhibit differences sufficiently large to indicate that a word is not a constant unit for the measurement of the rate of reading. For example, the rate score for Form 3 of the Courtis Silent Reading Test is 153 words per minute. For Brown's Silent Reading Test the rate is 182 words per minute. Similar differences are to be found in the

TABLE V. AVERAGE SCORES AND STANDARD DEVIATIONS FOR MEASURES OF RATE

Test	Grade IV				Grade VII			
	Form I		Form II		Form I		Form II	
	Av.	σ	Av.	σ	Av.	σ	Av.	σ
Monroe	79.5	24.9	94.9	21.1	104.0	30.9	140.7	24.8
Courtis	150.0	47.9	153.1	55.2	—	—	—	—
Brown	164.6	60.3	182.5	78.5	—	—	—	—
Starch	—	—	—	—	193.0	56.5	202.8	88.9
Reproduction	151.6	77.7	154.7	77.1	218.3	82.7	216.3	70.2
Cross-Out	75.3	28.2	84.0	22.2	111.7	30.1	133.8	34.3
Fordyce, Words	125.0	27.0	—	—	179.0	35.5	—	—
Pressey	—	—	—	—	24.0	1.5	23.5	1.7
Composition (Number of words written.)	—	—	—	—	218.6	85.5	—	—

seventh grade between Starch's Silent Reading Tests and the Experimental Reproduction Tests.

The rate scores for all of the tests are expressed in terms of words per minute. However, in the case of Monroe's Standardized Silent Reading Tests, The Cross-Out Silent Reading Tests, and the Pressey Silent Reading Test, the pupil does not do continuous reading. He must stop frequently to give responses. This, naturally, tends to reduce the rate scores. This is clearly shown in Table V. The rate scores for these tests are in most cases considerably less than rate scores in tests where the pupil does continuous reading. The difference is less marked in the seventh grade than in the fourth.

In Fordyce's Scale for Measuring Achievement in Reading, the pupil reads continuously, but the time allowance is such that a majority of the pupils complete the reading. Thus, they do not have an opportunity to give evidence of their rate of reading. This is the principal reason why the average rate scores for Fordyce's Tests are smaller than for the other tests in which the pupil does continuous reading.

The standard deviations also exhibit differences. Differences in the magnitude of the units would naturally affect the standard deviations as well as the averages. The standard deviation is also affected by the shape of the distribution. In a number of cases, the distribution of scores does not approximate the normal shape. This is, doubtless, one factor affecting the differences between the standard deviations.

Equivalence of duplicate forms The facts given in Tables IV and V indicate that the forms of these tests are not equivalent. In some cases an effort was made to construct the different forms so that they would be equivalent. This is true of Monroe's Standardized Silent Reading Tests. A study[22] planned to determine the degree of equivalence of these tests has indicated very definitely that they are not equivalent. The degree of non-equivalence revealed by that study is approximately that which is indicated here. The two forms of the Experimental Reproduction Tests, which were constructed without any preliminary study to determine their equivalence, appear to be as nearly equivalent as those of any other test in the list, as far as the rate is concerned. In the case of comprehension, there is considerable difference between the average scores. The two forms of the Cross-Out Tests were also constructed without much regard to equivalence and the average scores differ widely in most cases.

There is no published statement concerning the procedure followed by the authors of the other tests in order to secure equivalence of the duplicate forms. The average scores for the Courtis Silent Reading Test No. 2 do not differ widely. In fact, the two forms of this test appear to be the most nearly equivalent of any of the tests studied. The two Starch tests, No. 6 and No. 7, were not intended by the author to be equivalent forms. No. 7 (Form I) was intended to be more difficult, and lower average scores are, therefore, to be expected. This is what we find, except for the word-counting method of describing the reproductions. It is, however, obvious that it is difficult, or impossible, to construct duplicate forms which will be essentially equivalent, especially in the case of a small group of pupils. In addition to any lack of equivalence which may exist, the practise effect, due to one form being given after the other, would tend to produce differences between the average scores. The amount of this practise effect was not studied, since it was not pertinent to the major problem.

---

[22] Monroe, Walter S. Report of Division of Educational Tests for 1919-20. University of Illinois Bulletin, Vol. XVIII, No. 21, Page 19.

Relation of vocabulary to difficulty. In an effort to determine whether the vocabulary of a selection tends to determine its difficulty, the selections read by pupils in tests requiring reproduction were analyzed. All the words occurring in each selection were listed and the frequency of each one determined. The number of words in each selection not occurring in Ayres' list of one thousand words was also determined. In the case of the selections which formed duplicate tests, the vocabularies were compared, and the number of words common to the two selections was found. The results of this study are given in Table VI. For the Courtis Silent Reading Test, No. 2, 16 percent of the vocabulary in Form 1 and 19 percent of the vocabulary in Form 2 are not found in the Ayres' list. The number of different words, or the vocabulary, of Form 1 is 37 percent of the length of the selection. This means that, on the average, each word in the selection is used nearly three times. In the case of Form 3, the number of different words is 44 percent of the total number of words in the selection. The number of words common to the two selections is 15 percent of the average number of words in the two selections. These facts show that for these two forms of the Courtis Silent Reading Test, No. 2, the two selections are approximately equivalent with respect to the percent of words not found in the Ayres' list. Form 3 contains a slightly larger percent of words not in this list. Such words will, in general, be unusual words unless they are proper names. Form 3 has a rela-

TABLE VI. ANALYSIS OF SELECTIONS READ BY PUPILS IN SILENT READING TESTS REQUIRING REPRODUCTION

Test	Words not in Ayres' list.	Different words.	Words common to both selections.
Courtis I	.16	.37	.15
Courtis III	.19	.44	—
Starch, No. 6	.30	.55	.10
Starch, No. 7	.31	.59	—
Brown			
Long Slide	.13	.37	.13
Morning Adventure	.14	.35	—
Old English			
Heroes I	.19	.43	.13
II	.19	.44	—
The Strike at Shane's I	.19	.44	.13
II	.19	.52	—

tively larger vocabulary, and makes a greater demand upon a pupil's acquaintance with words. The percent of words which are common to the two selections is surprisingly small in view of the simple character of the material and of the fact that the two selections are considered equivalent in difficulty.

A comparison of the facts contained in Table VI with those in Tables IV and V indicates that the explanation for the non-equivalence of the two forms of the same test is not to be found in the vocabularies of the two selections in the respective tests. Evidently, the difficulty of a selection is determined by some factor other than the actual words used.

Formation of composite scores. The scores yielded by the different tests are expressed in terms of different scales. Therefore, it is necessary to reduce them to a common scale before combining them to form composite scores. The procedure adopted was to choose as a base the scale of Monroe's Standardized Silent Reading Test I, Form 1, for the fourth grade and the scale of Test II, Form 1, for the seventh grade. All other scores were reduced to the scale of these tests. The formula for reducing the scores obtained from one scale to equivalent scores on another scale is as follows:

$$S_1 = \frac{\sigma_1}{\sigma_2} S_2 + (Av_1 - \frac{\sigma_1}{\sigma_2} Av_2)$$

In this formula, S_2 is the obtained score on Form 2 and S_1 is the equivalent score expressed in terms of the scale of Form 1. Av_1 refers to the average of the scores obtained from Form 1; Av_2 refers to the average of the scores obtained from Form 2. The standard deviation of the distribution of the Form 1 scores is σ_1, and σ_2 is the standard deviation of the distribution of the Form 2 scores. This formula is based upon the usual assumption that corresponding deviations from averages are equal when expressed in terms of the standard deviation of the distribution; in other words, that

$$\frac{S_1 - Av_1}{\sigma_1} = \frac{S_2 - Av_2}{\sigma_2}$$

When this equation is solved for S_1 we obtain the formula as given above. The application of the above formula involves the determination of the numerical value of the ratio of $\frac{\sigma_1}{\sigma_2}$ by which the Form 2 score is to be multiplied and the determination of the numerical equivalent of the constant term of the formula (i. e., of the expression in parentheses). This latter numerical equivalent may be plus or minus. When it is positive it is to be added and when negative it is to be subtracted.

After the scores were reduced to the same scale composite scores were formed by calculating the averages of certain groups of scores. Composite AI is the average of Monroe, Form 1 (comprehension), Courtis, Form 1 (answers correct), and Reproduction, Form 1 (answers correct). (In the seventh grade the Courtis Test was not given and this composite score includes only the other two tests.) Composite AII is obtained from the second form of these tests.[23] Composite BI is the average of Brown's Silent Reading Test (both quality and quantity scores), and the Experimental Reproduction Tests (ideas and words). In the seventh grade, Starch's Silent Reading Test[24] (ideas and words) is used in the place of Brown's test. Composite CI is the average of Composite AI and Composite BI. Composite BII and CII were obtained in a corresponding way from the second forms of these tests. Composite I is obtained by combining all Form 1 scores. Composite II is obtained by combining all Form 2 scores.

Reliability. Since, with the exception of Fordyce's Scale for the Measurement of Achievement in Reading, two forms of each test were given, it is possible to compute measures of the extent to which equivalent scores were yielded by the different forms of a test. It is also possible to compute the *probable error of measurement* which is a measure of the magnitude of the departures of the obtained scores from the corresponding true scores.[25] These departures are the variable errors of measurement. No account is taken of the constant error of measurement in the following discussion. In the case of the tests for which the scoring is subjective, the computed reliability is greater than the true reliability for the reason that the averages of two independent scorings were used instead of the scores assigned by one person.[26]

Methods of determining reliability. In Tables VII and VIII, the reliability of these tests is described in terms of four quantities. (1) The coefficient of reliability is represented by the symbol, r_{12}, and is the coefficient of correlation between the two sets of scores yielded by the two forms of the test. (2) The index of reliability is represented by the symbol, $r_{1\infty}$. This quantity is the coefficient of correlation between one set of obtained scores and the set of corresponding true scores. The relation between the index of reliability and the coefficient of reliability is expressed by $r_{1t} = \sqrt{r_{12}}$. This formula was used in calculating the indices of reliability given in these two tables. (3) The probable error of measurement is represented by the symbol, P.E._M. This quantity was calculated by the formula, $\text{P.E.}_M = .6745\ \sigma\sqrt{1 - r_{12}}$.[27] The probable error of measurement (P. E._M) is a measure of the variable errors of measurement, or the differences between the obtained scores and the corresponding true scores. (4) The ratio of the probable error of measurement to the average of the scores from which it was calculated is represented by the symbol, $\frac{\text{P.E.}_M}{\text{Av.}}$. Table VII gives information concerning the reliability of rate scores and Table VIII, the corresponding information for comprehension scores. In case the test was scored by more than one method, the information is given for all methods of scoring.

---

[23] In the case of the Courtis Tests, Form 3 was used instead of Form 2.

[24] No. 7 is Form 1 and No. 6, Form 2.

[25] A true score is defined as the average of the scores yielded by a large number of duplicate forms of a test.

[26] See page 17 for the exact method used.

Probable error of r due to sampling. The coefficients of correlation, given in Tables VII and VIII and in the following tables, are subject to an error of sampling when interpreted with respect to the existence of relationship between the two sets of data from which they were derived. All of the correlations in the following tables are based on 80 cases in the fourth grade and 91 in the seventh

TABLE VII. MEASURES OF RELIABILITY, RATE

Test	Grade IV				Grade VII			
	r_{12}	r_{1t}	P.E._M	$\frac{\text{P.E.}_M}{\text{Av.}}$	r_{12}	r_{1t}	P.E._M	$\frac{\text{P.E.}_M}{\text{Av.}}$
Monroe I–II	.76	.87	11.3	.13	.63	.79	17.0	.11
Monroe I–III	.64	.80	13.6	.15	.55	.74	16.6	.11
Monroe II–III	.68	.82	11.8	.12	.69	.83	12.3	.09
Courtis	.85	.92	19.3	.13	—	—	—	—
Brown	.86	.93	26.0	.15	—	—	—	—
Starch	—	—	—	—	.62	.79	44.8	.23
Reproduction	.74	.86	39.5	.26	.45	.67	56.6	.26
Cross-Out	.68	.82	14.4	.18	.76	.87	15.8	.13
Pressey	—	—	—	—	.50	.71	1.1	.05

---

[27] For explanation of this formula and the method of application see page 24.

TABLE VIII. MEASURES OF RELIABILITY, COMPREHENSION

Test	Grade IV				Grade VII			
	r_{12}	r_{1t}	P.E.$_{\mathrm{M}}$	$\frac{\text{P.E.}_{\mathrm{M}}}{\text{Av.}}$	r_{12}	r_{1t}	P.E.$_{\mathrm{M}}$	$\frac{\text{P.E.}_{\mathrm{M}}}{\text{Av.}}$
Monroe I–II	.73	.85	3.0	.21	.69	.83	5.2	.20
Monroe I–III	.65	.81	3.4	.22	.60	.77	5.6	.20
Monroe II–III	.78	.88	2.5	.15	.61	.78	5.2	.17
Courtis, Index	.58	.76	9.9	.12	—	—	—	—
Courtis, No. of Questions	.80	.89	4.6	.11	—	—	—	—
Courtis, No. of Questions Correct	.69	.83	5.7	.16	—	—	—	—
Brown, Quantity	.36	.60	14.2	.37	—	—	—	—
Brown, Quality	.19	.44	13.1	.54	—	—	—	—
Brown, Average	.32	.57	16.6	.55	—	—	—	—
Brown, Efficiency	.47	.69	20.1	.43	—	—	—	—
Brown, Words	.63	.79	13.7	.27	—	—	—	—
Brown, Ideas	.48	.69	6.6	.28	—	—	—	—
Starch, Words	—	—	—	—	.77	.88	9.7	.25
Starch, Ideas	—	—	—	—	.72	.85	4.8	.27
Reproduction, Questions	.48	.69	1.9	.20	.60	.78	1.8	.20
Reproduction, Ideas	.54	.73	7.1	.37	.72	.85	15.1	.15
Reproduction, Words	.50	.71	19.5	.41	.87	.93	24.6	.19
Reproduction, Questions and Words	.44	.63	4.0	.30	.64	.80	4.6	.20
Cross-Out C−W	.62	.79	3.4	.47	.67	.82	4.6	.26
Cross-Out $\frac{C-W}{C+O}$	.52	.72	19.9	.47	.52	.72	14.7	.21
Pressey	—	—	—	—	.65	.81	1.9	.14
Memory, Ideas	.35	.59	4.3	.18	.56	.75	3.3	.10
Memory, Words	.40	.63	12.8	.17	.34	.58	13.3	.13

grade. In order to economize space, we give in Table IX probable errors due to sampling for various values of r. Most of the coefficients of correlation appearing in these tables are sufficiently large in comparison with the probable error due to sampling that they may be interpreted as indicating the existence of a distinct positive relationship. We are, however, more interested in securing a measure of the departure from perfect correlation. Hence, the probable error of measurement (P.E.$_{\mathrm{M}}$) is a much better index of the degree of reliability of a test than either r_{12} or r_{1t}.

Reliability of the tests studied. Brown's Silent Reading Test, when scored in the way which he recommends, is the least reliable. The ratio of the probable error of measurement to the average is .54 for the quality and .55 for the average of quantity and quality.

TABLE IX. PROBABLE ERRORS OF THE COEFFICIENT OF CORRELATION (r_{12}) DUE TO USING A LIMITED NUMBER OF CASES*

r_{12}	P. E.
.1	.0798
.2	.0774
.3	.0734
.4	.0677
.5	.0605
.6	.0516
.65	.0466
.70	.0411
.75	.0353
.80	.0290
.85	.0224
.90	.0153
.95	.0079

*80 in Grade IV and 91 in Grade VII.

The "efficiency score" has a ratio of .43. The scoring of this test by means of either the idea-counting method or the word-counting method results in scores that are more reliable. Considering both rate and comprehension, the most reliable test is the Courtis Silent Reading Test, No. 2. For rate, the index of reliability is .92 and the ratio of the probable error of measurement to the average is .13. Three comprehension scores are used in connection with this test. The number of questions answered is shown to be the most reliable.

The probable error of measurement and the ratio of the probable error of measurement to the average score indicate a degree of reliability for the rate scores yielded by Monroe's Standardized Silent Reading Tests which is surprisingly high, considering the character of the tests. This is particularly true in the seventh grade. With the exception of Pressey's Silent Reading Test, they are the most reliable. In the fourth grade, the reliability is exceeded only by Courtis's Silent Reading Test, No. 2. In Monroe's Standardized Silent Reading Tests a pupil does not read continuously but is forced to stop at the end of each exercise and answer a question. According to the rules for scoring these tests, a pupil receives no credit for an exercise unless he has completed his reading of it to the extent of recording his answer. The increments added to a rate score for doing additional exercises are relatively large, particularly in Test II. Thus, a pupil who has failed only in recording his answer to an exercise receives a score which does not indicate his rate of reading. His score is the same as that of the pupil who has just barely completed the preceding exercise. In all of the other tests, with the exception of Pressey's Silent Reading Test, the pupil's rate score represents the actual amount read. In view of these facts, it is surprising to find that Monroe's Standardized Silent Reading Tests yield rate scores which have such a high degree of reliability. The figures which are given may be affected somewhat by the fact that these tests proved too short and a considerable number of pupils made perfect scores.

In general, the degree of reliability is higher in the seventh grade than in the fourth. Exact comparisons cannot be made because identical tests were not given in the two grades; but where similar tests were given the results for the seventh grade show a distinctly higher degree of reliability. This may be due to a superiority in the tests for the seventh grade or it may be due to the fact that the increased maturity of the pupils causes them to be less variable in their performances.

The degree of unreliability shown in Tables VII and VIII is distressingly high. As we have indicated, the ratio of the probable error of measurement to the average probably furnishes the most significant statement of the degree of unreliability. Brown's Test, scored by any method, appears to be so highly unreliable that it should be rejected. In interpreting the figures in Table VIII it should be borne in mind that the actual degree of unreliability is somewhat larger than that indicated because the element of subjectivity in scoring has been largely eliminated. It appears that individual scores yielded by these tests are very imperfect measures of reading ability. However, the variable errors involved do not affect, to the same degree, the scores of classes or larger groups. Although the scores yielded by these tests must be considered as having only a very limited significance in the case of individual pupils, they are much more significant for groups of pupils.

Both the Experimental Reproduction Tests and the Cross-Out Tests were merely experimental. The reproduction tests were intentionally so. It was desired to ascertain whether a crude reproduction test, such as might be constructed by a teacher and administered directly from a supplementary reader, would yield results as reliable as tests more carefully constructed and more conveniently arranged. These tests are shown to be among the least reliable, with the exception of Brown's Silent Reading Test. This is to be expected; but the difference in reliability, particularly in the seventh grade, is not marked. In fact, the Experimental Reproduction Tests exhibit a relatively high degree of reliability in the measurement of comprehension. Thus, the reliability of a crude test of this type is only slightly less than that of tests whose construction was more refined.

Discrimination. The distributions of the rate scores yielded by the different tests indicate that certain tests fail to yield scores which discriminate between a number of pupils with respect to rate of reading. Form 3 of Monroe's Standardized Silent Reading Tests I and II is clearly too short. In the seventh grade 58 percent of the pupils and in the fourth grade 27 percent completed the test. All such pupils received the maximum rate score. The distributions for Forms 1 and 2 of this test contain no such extreme deviations from the normal shape, although Form 2 of Test I and Form 1 of Test II cannot be said to approximate closely the normal distribution.

The Cross-Out Tests yield distributions which exhibit many irregularities and which cannot be said to do more than suggest the normal distribution. As was to be expected, a large percent of the pupils completed the reading of the selection in the case of the Fordyce test. Forty-nine percent of the pupils in the fourth grade and 29 percent in the seventh grade received the maximum rate score. The Pressey Test proved too short for the time allowed. Seventy-six percent completed Form 1 and 56 percent completed Form 2. The Courtis, Brown, Starch, and Experimental Reproduction Tests yielded rate scores which formed distributions closely approximating the normal shape. A few irregularities were exhibited by the Experimental Reproduction Tests and by Brown's test.

As judged by the shape of the distribution of the rate scores, the Courtis Silent Reading Test, No. 2, exhibits the least lack of discrimination. The Cross-Out, Pressey, Fordyce, and Form 3 of Monroe's tests exhibit such great departures from the normal distribution that they must, obviously, fail to discriminate properly with respect to the rate of reading for a considerable number of pupils.

In the case of comprehension, the distributions of scores for Monroe's Standardized Silent Reading Tests closely approximate the normal. The third form appears to have been a little too easy; but, in other respects, the irregularities exhibited by the distributions cannot be considered to indicate a serious lack of discrimination. The index of comprehension for the Courtis Silent Reading Test, No. 2, fails to discriminate properly between a number of pupils. Both the number of questions answered and the number of questions answered correctly approach more nearly the normal distribution.

TABLE X. CORRELATIONS WITH TEACHER RATING

Test	Rate		Comprehension	
	Grade IV	Grade VII	Grade IV	Grade VII
Monroe I	.38	.25	.60	.32
Monroe II	.34	.35	.64	.50
Monroe III	.43	.26	.63	.39
Courtis I Index			.29	
Courtis I Questions			.29	
Courtis I Questions Correct			.41	
Courtis I Words per minute	.55			
Courtis III Index			.45	
Courtis III Questions			.38	
Courtis III Questions Correct			.51	
Courtis III Words per minute	.51			
Brown I Quantity			.30	
Brown I Quality			.29	
Brown I Average			.34	
Brown I Efficiency			.58	
Brown I Words			.64	
Brown I Ideas			.55	
Brown I Words per minute	.53			
Brown II Quantity			.01	
Brown II Quality			—.05	
Brown II Average			.05	
Brown II Efficiency			.36	
Brown II Words			.48	
Brown II Ideas			.45	
Brown II Words per minute	.42			
Starch I Words				.46
Starch I Ideas				.46
Starch I Words per minute		.19		
Starch II Words				.51
Starch II Ideas				.49
Starch II Words per minute		.41		
Reproduction I Questions			.42	.34
Reproduction I Ideas			.31	.47
Reproduction I Words			.35	.23
Reproduction I Words per minute	.36	.26		
Reproduction II Questions			.36	.51
Reproduction II Ideas			.51	.49
Reproduction II Words			.43	.46
Reproduction II Words per minute	.32	.16		
Cross-Out I C−W			.39	.21
Cross-Out I $\frac{C-W}{C+O}$			.39	.27
Cross-Out I Words per minute	.36	.40		
Cross-Out II C−W			.56	.34
Cross-Out II $\frac{C-W}{C+O}$			.46	.20
Cross-Out II Words per minute	.43	.36		
Fordyce	.49	.13	.24	.39
Pressey I		.29		.46
Pressey II		.08		.37
Composite AI			.58	.40
Composite AII			.60	.55
Composite BI			.44	.53
Composite BII			.34	.58
Composite CI			.59	.51
Composite CII			.56	.63
Composite I	.55	.45		.58
Composite II	.51	.36		.58

This is particularly true of the latter. The distributions for the Brown, Starch, and Experimental Reproduction Tests exhibit many irregularities; but there is in all cases a distinct resemblance to the normal distribution. A few of the distributions approach very closely the normal one. Others contain rather marked departures from it. In the case of Brown's test, the distributions for the quality scores exhibit greater departures thant he distributions for the quantity scores.

Comparison with teachers' ratings. All scores, both rate and comprehension, were correlated with the ratings in silent reading given by the teacher. The coefficients of correlation were calculated, also, for certain composite scores. These coefficients of correlation are given in Table X. With the exception of one coeffiient for the second form of Brown's test, all coefficients are positive and in general sufficiently large to indicate a distinct positive relationship between the test scores and the teachers' ratings. Rate of reading correlates more highly with the teachers' rating in the fourth grade than in the seventh. For rate, the average of the coefficients, not including the composite scores, is 43 in the fourth grade and 26 in the seventh. The average of the coefficients for comprehension, not including the composite scores, is 40 in the fourth grade and 44 in the seventh.

In the fourth grade, comprehension, as measured by Monroe's Standardized Silent Reading Tests, correlates most highly with the teachers' ratings. In fact, the coefficients for the three forms of this test equal or exceed all of those for the composite scores. In the seventh grade this test does not exhibit as high correlations with teachers' ratings. Neither do its rate scores correlate as highly with teachers' ratings as the rate scores yielded by some other tests. It is interesting to note that the correlation between the second form of Brown's Test for "quantity of reproduction" and "quality of reproduction" is essentially zero. For Form 1 the correlations for these two scores are lower than the correlations for any other scores. This suggests that Brown's method for scoring his test is undesirable. The correlations of the composite scores with teachers' ratings indicate that, in the fourth grade, teachers judge silent reading ability more on the basis of the pupils' ability to answer questions than of their ability to reproduce. In the seventh grade, the teachers give greater weight to the pupils' ability to reproduce or to tell what has been read.

Correlation of comprehension with memory. In those tests which require the pupil to answer questions from memory or to

TABLE XI. CORRELATION OF COMPREHENSION WITH MEMORY

Test	Grade IV				Grade VII			
	Ideas		Words		Ideas		Words	
	I	II	I	II	I	II	I	II
Brown I Quantity	.32	.21	.39	.27	—	—	—	—
Brown II Quantity	.27	.20	.23	.20	—	—	—	—
Brown I Quality	.36	.16	.36	.16	—	—	—	—
Brown II Quality	.19	.11	.14	.08	—	—	—	—
Starch I Ideas	—	—	—	—	.36	.27	.32	.31
Starch II Ideas	—	—	—	—	.39	.24	.34	.25
Starch I Words	—	—	—	—	.37	.28	.33	.47
Starch II Words	—	—	—	—	.24	.33	.45	.34
Reproduction I Questions	.46	.35	.46	.33	.43	.34	.33	.26
Reproduction II Questions	.27	.38	.30	.36	.51	.27	.44	.20
Reproduction I Ideas	.42	.41	.46	.41	.53	.43	.55	.36
Reproduction II Ideas	.39	.24	.35	.28	.56	.40	.50	.35
Reproduction I Words	.37	.41	.35	.41	.45	.35	.53	.33
Reproduction II Words	.41	.28	.38	.27	.54	.40	.53	.39
Monroe I	.31	.33	.28	.27	.49	.47	.34	.35
Monroe II	.25	.33	.24	.25	.52	.48	.45	.24
Monroe III	.31	.35	.35	.38	.37	.33	.40	.26
Maximum	.46	.41	.46	.41	.56	.48	.55	.47
Minimum	.19	.11	.14	.08	.24	.24	.32	.20
Average	.34	.29	.33	.28	.44	.36	.42	.32

TABLE XII. CORRECTED COEFFICIENTS OF CORRELATION OF COMPREHENSION WITH MEMORY

Test	Grade IV		Grade VII	
	Ideas	Words	Ideas	Words
Brown Quantity	.67	.66	—	—
Brown Quality	.68	.54	—	—
Starch Ideas	—	—	.51	.66
Starch Words	—	—	.39	.90
Reproduction Questions	.75	.72	.72	.75
Reproduction Ideas	.97	.86	.77	.86
Reproduction Words	.98	.88	.62	.77
Monroe I–II	.57	.47	.80	.82
Monroe I–III	.68	.61	.72	.83
Monroe II–III	.61	.58	.71	.75

reproduce the passage read, it would seem that a pupil's ability to remember would materially affect his comprehension score. In order to ascertain the extent to which ability to remember does affect the comprehension score yielded by such tests, the pupils were given the memory test[28] described on page 7. In this test a selection was read to the pupils and they were asked to reproduce the story from memory. The coefficients of correlation between the memory scores and the comprehension scores for silent reading tests are given in Table XI. It is significant that none of these coefficients are large. The first three tests listed in this table require the pupil to give his performances from memory. Monroe's Standardized Silent Reading Tests do not appear to make any considerable demand upon the pupil's memory; he has the passage before him and can read it and re-read it if he desires. If any memory is involved it is immediate in character. It is significant that the coefficients of correlation for this test closely approximate those for other tests.

Corrected coefficients of correlation. The measures yielded by these tests involve variable errors. It has been shown in our consideration of the reliability of these tests that these errors are relatively large for the reproduction tests. The presence of these variable errors tends to reduce the coefficients of correlation, and it is possible that the coefficients of correlation given in Table XI do not represent the true relation between comprehension and memory.

When two forms of both tests have been given to the same pupils it is possible to compute a corrected coefficient of correlation which is free from the effect of the variable errors of measurement. This has been done by means of the following formula:[29]

$$r_{pq} = \frac{\sqrt{(r_{p_1q_2})\ (r_{p_2q_1})}}{\sqrt{(r_{p_1p_2})\ (r_{q_1q_2})}}$$

r_{pq} here indicates the true correlation between two series of measures, $_p$ and $_q$, of the facts A and B.

p_1 and p_2 are two independent measures of A.

q_1 and q_2 are two independent measures of B.

$r_{p_1q_2}$ is the correlation obtained from the first measure of A and the second measure of B.

$r_{p_2q_1}$ is the correlation obtained from the second measure of A and the first measure of B.

---

[28] It is assumed that this test measures ability to remember.

[29] Thorndike, E. L. "An Introduction to Mental and Social Measurements." New York. Teachers College, Columbia University, 1916. Page 179.

$r_{p_1p_2}$ is the correlation between the two measures of A.
$r_{q_1q_2}$ is the correlation between the two measures of B.

In applying this formula the factors of the numerator are obtained from Table XI. For example, in calculating the corrected coefficient of correlation for Brown's Silent Reading Test with memory, $r_{p_1q_2}$ is the coefficient of correlation of Brown I with Memory II. This is given as .21. The coefficient of correlation of Brown II with Memory I, is $r_{p_2q_1}$. This is given as .27. The factors of the denominator are the reliability coefficients of the two tests. These are to be found in Table VIII. They are .36 for Brown's Silent Reading Tests and .35 for the Memory Tests. Substituting these values in the formula,

$$
\begin{aligned}
r_{pq} &= \frac{\sqrt{.21 \text{ x } .27}}{\sqrt{.36 \text{ x } .35}} \\
&= \sqrt{.45} \\
&= .67
\end{aligned}
$$

This is the first entry of the first column of Table XII.

A study of the corrected coefficients given in Table XII indicates that, in the case of the Experimental Reproduction Tests in the fourth grade, the correlation between Memory and the scores based upon the pupil's reproduction is very high. For ideas it is .97. For words it is .88. For Brown's Silent Reading Tests the correlation is not as high. In fact, it closely approximates that for Monroe's Standardized Silent Reading Tests. In the seventh grade the correlation of Memory with Monroe's Standardized Silent Reading Tests is higher than that for either Starch or the Experimental Reproduction Tests, although the difference is not marked in the case of the latter. It, therefore, appears that in the seventh grade memory is not a major factor in determining the comprehension scores of tests which require reproduction unless it is also the determining factor in the case of tests which do not appear to involve memory. The statement which has been made with reference to reproduction tests, that they measure the ability to read *and remember*, does not appear to be justified by the facts which are presented here.

Correlation of comprehension with vocabulary. In Table XIII, we give the coefficients of correlation between the comprehension scores and the scores obtained from the vocabulary test. In the fourth grade most of the coefficients are negative, but all of them cluster closely around zero. This means that, measured by the tests used,

TABLE XIII. COEFFICIENTS OF CORRELATION BETWEEN VOCABULARY AND COMPREHENSION

Test	Grade IV	Grade VII
Monroe I	.02	.22
Monroe II	−.03	.22
Monroe III	−.02	.13
Courtis I Index	−.20	—
Courtis I No. of Questions	.19	—
Courtis I Questions Correct	.10	—
Courtis III Index	−.20	—
Courtis III No. of Questions	.06	—
Courtis III Questions Correct	.15	—
Starch I Words	—	.31
Starch I Ideas	—	.32
Starch II Words	—	.29
Starch II Ideas	—	.22
Brown I Quantity	−.11	—
Brown I Quality	−.12	—
Brown I Average	.14	—
Brown I Words	.01	—
Brown I Ideas	−.04	—
Brown II Quantity	−.23	—
Brown II Quality	−.21	—
Brown II Average	−.16	—
Brown II Words	−.15	—
Brown II Ideas	−.19	—
Reproduction I Questions	−.15	.14
Reproduction I Ideas	−.10	.17
Reproduction I Words	−.09	.13
Reproduction I Questions	.12	.19
Reproduction I Ideas	−.04	.24
Reproduction I Words	−.04	.26
Cross-Out I C−W	−.07	.18
Cross-Out I $\frac{C-W}{C+O}$	−.05	.08
Cross-Out II C−W	−.09	.16
Cross-Out II $\frac{C-W}{C+O}$	−.02	.01
Fordyce	.04	.13
Pressey I	—	.21
Pressey II	—	.00
Composite AI	−.02	.23
Composite AII	.01	.20
Composite BI	−.08	.32
Composite BII	−.20	.28
Composite CI	−.05	.26
Composite CII	−.13	.25
Composite I	—	.30
Composite II	—	.21

there is no relation between a pupil's vocabulary and his ability to read. It is, of course, obvious that, in order to read, a pupil must be acquainted with words. It is, therefore, impossible to believe that vocabulary is not a factor in the reading process. The facts presented here probably mean that, in the fourth grade, vocabulary is not a determining factor and the pupil's ability to read depends primarily upon abilities other than the extent of his acquaintance with words. In the seventh grade the coefficients are all positive but none of them are large. This probably means that, in the seventh grade, vocabulary is a minor factor in determining the pupil's comprehension. It is, of course, possible that the vocabulary test used does not measure the extent of a pupil's acquaintance with words.

Correlation of cancellation scores with measures of rate of reading. In Table XIV, the coefficients of correlation for the scores yielded by the Cancellation Test with measures of rate of silent reading are given. With few exceptions, these coefficients are positive but small. In general, they are slightly smaller in the seventh grade than in the fourth grade. In most cases, there does not seem to be any marked relationship between ability to do the Cancellation Test and the rate of silent reading. One might expect a distinct positive relationship between the Cross-Out Silent Reading Tests and the Pressey Silent Reading Tests. It does, however, appear that the relationship which exists with respect to these tests is greater than that which exists for Monroe's Silent Reading Tests.

The table also includes coefficients of correlation for the scores yielded by the Cancellation Test with the comprehension scores yielded by the Cross-Out Tests. The coefficients are, likewise, small, two of them being slightly negative. It appears, therefore, that the ability to strike out letters from words is not related to the ability called for by the Cross-Out Tests.

Correlation of comprehension with written composition. Another measure of a pupil's vocabulary is secured from his written composition. The pupils in the seventh grade were asked to write a composition on an exciting experience. (See page 10.) In Table XV, we give the coefficients of correlation between measures of comprehension and two measures of these written compositions, the number of words written and the story value. The number of words which a pupil writes in such an exercise is, undoubtedly, an index of his writing vocabulary. It is, of course, possible that his writing

TABLE XIV. CORRELATION OF CANCELLATION SCORES WITH MEASURES OF RATE OF READING AND WITH THE CROSS-OUT TESTS

Test	Grade IV		Grade VII	
	Cancellation		Cancellation	
	I*	II*	I	II
Monroe I	.28	.14	.07	.20
Monroe II	.26	.15	.10	.18
Monroe III	.23	.20	−.03	.22
Courtis I	.12	.15		
Courtis III	.20	.14		
Brown I	.25	.16		
Brown II	.07	.08		
Starch I			−.06	−.01
Starch II			−.15	.03
Reproduction I	.22	.13	.05	.25
Reproduction II	.13	.08	.10	.22
Cross-Out I	.20	.15	.18	.25
Cross-Out II	.23	.13	.14	.33
Fordyce, No. of Words	.30	.12	−.08	.11
Pressey I			−.03	.11
Pressey II			.06	.15
Cross-Out I $C-W$	.08	.01	.21	.18
Cross-Out I $\frac{C-W}{C+O}$	.02	−.03	.11	.05
Cross-Out II $C-W$	.17	.06	.16	.17
Cross-Out II $\frac{C-W}{C+O}$	.14	.03	.11	−.01

*In Cancellation Test I, the words containing both "a" and "t" were marked; in Test II, those containing both "e" and "r."

vocabulary and his reading vocabulary are not closely related. The coefficients of correlation, in Table XV, show that there is little or no relation existing between measures of comprehension and the number of words which were written in these compositions. Even in the case of comprehension scores based upon the number of words and the number of ideas contained in reproductions, the coefficients of correlation fail to indicate the existence of any marked relationship. In fact, the coefficients of correlation for measures of comprehension gained through reproduction are lower, in most cases, than the coefficients of correlation of the number of words written with the comprehension scores derived from Monroe's Standardized Silent Reading Tests.

A higher degree of correlation is indicated between the "story value" and the measures of comprehension. Some of the coefficients of correlation are sufficiently large to indicate a distinct positive relationship between these two traits. It is not unlikely that this relationship can be explained in terms of a common general factor, such as general intelligence.

Inter-correlation between tests. Since in each grade all of the tests were given to the same pupils, it is possible to calculate the coefficients of correlation between scores yielded by the different tests. These are given in the appendix. The magnitude of the coefficients of correlation is influenced by the reliability of the scores and, therefore, does not truthfully reflect the relationship which exists between the scores yielded by the different tests. In order to secure more accurate indices of the relationship existing between traits measured by the different tests, the corrected coefficients of correlation have been calculated by means of the formula given on page 41. Since the factors of both numerator and denominator of the formula are square roots, it is impossible to calculate corrected coefficients when one of the raw coefficients is negative. This

TABLE XV. CORRELATION OF COMPREHENSION WITH WRITTEN COMPOSITION, SEVENTH GRADE, 90 PUPILS

Test	Number of words written	Story value
Monroe I	.18	.29
Monroe II	.23	.33
Monroe III	.24	.31
Starch I Words	.10	.31
Starch I Ideas	.07	.28
Starch II Words	.14	.36
Starch II Ideas	.09	.33
Reproduction I Questions	.12	.24
Reproduction I Ideas	.22	.14
Reproduction I Words	.22	.18
Reproduction II Questions	−.07	.11
Reproduction II Ideas	.26	.37
Reproduction II Words	.28	.43
Cross-Out I C−W	.13	.23
Cross-Out I $\frac{C-W}{C+O}$	.09	.11
Cross-Out II C−W	.16	.06
Cross-Out II $\frac{C-W}{C+O}$	.04	.11
Fordyce Percent	.12	.12
Pressey I	.10	.29
Pressey II	.05	.18

accounts for the fact that certain corrected coefficients are not given in Tables XVI and XVII. It will be noted in these tables that, occasionally, a coefficient greater than 1.00 is given. This is due to chance errors in the raw coefficients of correlation which, in turn, are due to the fact that a sample of the total population was used in calculating them. The corrected coefficients are, in general, larger than the corresponding raw coefficients.

Table XVI gives the corrected coefficients for the comprehension scores. A significant characteristic of this table is the variation in the degree of intercorrelation between the tests. For example, Monroe's Standardized Silent Reading Test I correlates very highly with the number of questions answered correctly on the Courtis Silent Reading Test, No. 2. It correlates less highly with the other two scores of this test. The degree of its correlation with the other tests is moderately low. It is significant that the corrected coefficients of correlation between the two tests requiring reproduction are not higher. For example, the highest coefficient of correlation between Brown's test and the Experimental Reproduction Test I is .79. The lowest is .26. The corrected coefficient of correlation between the scores obtained by the word-counting method is .33; for the idea-counting method the coefficient of correlation is .62. The highest correlation between Brown's test and the Experimental Reproduction Test I is for the number of questions answered correctly. In the seventh grade, the corrected coefficients of correlation between the question scores yielded by the Experimental Reproduction Test II and Starch's Silent Reading Test are as high as those obtained from the reproductions. Both Starch's test and the Experimental Reproduction Test correlate nearly as highly with Monroe's Standardized Silent Reading Test as with each other. A number of the coefficients of correlation for the Cross-Out Test are relatively high. It correlates most highly with Monroe's Standardized Silent Reading Test. In general, the coefficients are higher for the scores obtained by C—W than for $\frac{C-W}{C+O}$. The former is probably the better plan of scoring.

Table XVI appears to bear out the usual assumption that different silent reading tests measure different phases of silent reading ability. It is very obvious, in a number of cases, that the same traits are not measured by different tests. However, it should be noted that these differences exist for tests that are similar in structure as well as for tests which possess marked differences in struc-

TABLE XVI. CORRECTED COEFFICIENTS OF CORRELATION OF MEASURES OF COMPREHENSION OF READING

TESTS	Grade IV			Grade VII			Grade IV Courtis			Grade IV Brown					Grade VII Starch		Grade IV Reprod.			Grade VII Reprod.			Grade IV Cross-Out		Grade VII Cross-Out		Grade VII Pressey
	I–II	I–III	II–III	I–II	I–III	II–III	Index of Comp.	No. of Questions	Questions Correct	Quantity	Quality	Average	Words	Ideas	Words	Ideas	Questions	Ideas	Words	Questions	Ideas	Words	C−W	$\frac{C-W}{C+O}$	C−W	$\frac{C-W}{C+O}$	
Monroe I-II							.57	.66	.92				.65	.59	.72	.72	.55	.47	.44	.65	.49	.39	.69	.44	.87	.64	.69
I-III							.62	.72	.97				.73	.67	.68	.59	.63	.59	.56	.58	.48	.38	.79	.33	.84	.72	.66
II-III							.72	.57	.89	.00			.73	.70	.75	.74	.66	.48	.48	.62	.52	.46	.75	.52	.91	.65	.83
Courtis																											
Index of comprehension	.57	.62	.72							.40	.55	.38	.50	.49			.59	.33	.38				.63	.55			
No. of Questions	.66	.72	.57								.15		.57	.56			.50	.36	.35				.39	.15			
Questions Correct	.92	.97	.89							.18			.79	.73			.76	.47	.52				.66	.41			
Brown																											
Quantity			.00				.40		.18								.44	.48	.59				.28	.57			
Quality							.55	.15									.61	.39	.26				.22	.61			
Average							.38										.43	.55	.47				.19	.55			
Words	.65	.73	.73				.50	.57	.79								.79	.69	.33				.81	.71			
Ideas	.59	.67	.70				.49	.56	.73								.72	.62	.68				.71	.71			
Starch																											
Words				.72	.68	.75														.57	.51	.54			.56	.52	.54
Ideas				.72	.59	.74														.51	.57	.51			.63	.50	.49
Reproduction																											
Questions	.55	.63	.66	.65	.58	.62	.59	.50	.76	.44	.61	.43	.79	.72	.57	.51							.67	.52	.58	.65	.73
Ideas	.47	.59	.48	.49	.48	.52	.33	.36	.47	.48	.39	.55	.69	.62	.49	.58							.27	.33	.53	.48	.63
Words	.44	.56	.48	.39	.38	.46	.38.	.35	.52	.59	.26	.47	.33	.68	.54	.51							.47	.42	.46	.44	.53
Cross-Out																											
C−W	.69	.79	.75	.87	.84	.91	.63	.39	.66	.28	.22	.19	.81	.71	.56	.63	.67	.27	.47	.58	.53	.46					.76
$\frac{C-W}{C+O}$	.44	.33	.52	.64	.72	.65	.55	.15	.41	.57	.61	.55	.71	.71	.52	.50	.52	.33	.42	.65	.48	.44					.79
Pressey				.69	.66	.83									.54	.49				.73	.63	.53			.76	.79	
Composite																											
A	.97	.97	.97	.96	.90	.97	.53	.79	1.06				.79	.78	.71	.69	.63	.60	.53	.91	.51	.48	.69	.37	.79	.62	.78
B	.40	.42	.44	.52	.49	.71	.42	.68	.38	.79	.80	.84	.68	.87	.89	.83	.77	.90	.95	.57	.88	.78	.56	.56	.53	.60	.57
C	.80	.86	.82	.75	.74	.84	.52	.62	.80	.43	.88	.48	.80	.91	.88	.96	.89	.83	.83	.84	.82	.72	.73	.54	.76	.66	.85
General				.84	.83	.89									.81	.80				.82	.77	.64			.84	.76	.87

TABLE XVII. CORRECTED COEFFICIENTS OF CORRELATION OF MEASURES OF RATE OF READING

	Grade IV Monroe			Grade VII Monroe			Grade IV Courtis	Grade IV Brown	Grade VII Starch	Grade IV Reprod.	Grade VII Reprod.	Grade IV Cross Out	Grade VII Cross Out	Grade VII Pressey	Grade IV Comp.	Grade VII Comp.	Grade VII Comp. A
	I–II	I–III	II–III	I–II	I–III	II–III											
Monroe I–II							.75	.69	.66	.74	.83	.91	.72	.81	.90	.94	.73
I–III							.82	.73	.64	.84	.87	.94	.61	.85	.94	.95	.77
II–III							.72	.73	.62	.66	.62	.95	.66	.84	.90	.91	.65
Courtis	.75	.82	.72					.84		.90		.76			.94		
Brown	.69	.73	.73				.84			.91		.68			.89		
Starch				.66	.64	.62					.72		.57	.57		.85	.86
Reproduction	.74	.84	.66	.83	.87	.62	.90	.91	.72			.73	.67	.41	.90	.95	1.04
Cross-Out	.91	.94	.95	.72	.61	.66	.76	.68	.57	.73	.67			.47	.83	.83	.64
Pressey				.81	.85	.84			.57		.41		.47			.57	.60
Composite	.90	.94	.90	.94	.95	.91	.94	.89	.85	.90	.95	.83	.83	.57			.93
Composite A				.73	.77	.65			.86		1.04		.64	.60		.93	

ture. In fact, the variations in these corrected coefficients of correlation are so erratic that one is inclined to be skeptical of any conclusions which may be drawn from them with reference to the functions of the different tests.

The corrected coefficients for the rate scores are given in Table XVII. These are, in general, higher than those for comprehension. In general, the correlation between tests in which the pupil reads continuously is higher than between one test in which the pupil reads continuously and another in which his reading is not continuous. However, the correlation between Monroe's Standardized Silent Reading Test I and the Cross-Out Test, in the fourth grade, is as high as that for any of the other tests. The fact that some of the tests were too short and failed to discriminate between a considerable number of pupils probably accounts for the fact that a number of coefficients of correlation are not higher. An examination of this table indicates that the rate score secured by means of Monroe's Standardized Silent Reading Tests is a true measure of the pupil's rate of reading.

Correlation of single tests with composites. In Tables XVI and XVII, the corrected coefficients of correlation for each test with certain composite scores are given. These, in general, are larger than the coefficients of correlation between single tests. In the fourth grade, composite A for comprehension is the average of Monroe, comprehension, Courtis, answers correct, and Reproduction, answers to questions. In the seventh grade, the Courtis test was not given and this composite includes only the other two tests. Composite B for comprehension is the average of the comprehension scores derived from reproductions. In the case of Brown's Silent Reading Tests, both quality and quantity are used. In the other cases, the scores obtained by both the idea-counting method and the word-counting method are used. Composite C is the average of composite A and composite B. The general composite is formed by combining all of the scores obtained.

Monroe's Standardized Silent Reading Tests are shown to correlate very highly with composite A. The correlation with composite B is very much less, as might be expected. The rate scores derived from this test also correlate very highly with the general composite scores. In fact, with the exception of Pressey's test, the correlation of single tests with the composite scores is very high. It appears, therefore, that each of the tests yields rate scores which may be accepted as correlating very highly with the true rate of silent reading. The scores derived from the Experimental Reproduction Tests in the fourth grade correlate more highly with composite B than those derived from Brown's Silent Reading Test. In the seventh grade, the correlations between Starch's test and composite B are slightly higher than those for the Experimental Reproduction Tests. It appears, however, that the Experimental Reproduction Tests yield approximately as valid measurements of ability to comprehend as are secured by means of the other tests which, presumably, have been devised with greater care.

SUMMARY OF CONCLUSIONS.

1. The scoring of reproductions is so highly subjective that a silent reading test requiring reproduction of material read cannot be considered satisfactory.

2. Brown's Silent Reading Test is very unreliable for both comprehension and rate. This is true, even when the average of two independent scores is used as a measure of comprehension.

3. The correlation between scores yielded by the memory test and comprehension scores based upon reproductions is only slightly higher than that existing between the scores derived from the memory test and the comprehension scores yielded by Monroe's Standardized Silent Reading Test. This makes doubtful the usual assumption that measures of comprehension based upon reproductions are affected by the pupil's ability to remember.

4. Correlation between extent of vocabulary and ability to read is surprisingly low. There is little, if any, relation between these two abilities.

5. The intercorrelations between tests indicate that different tests measure slightly different traits; but it is surprising to find, in a few instances, a high degree of correlation existing between scores yielded by tests which exhibit marked differences in structure.

6. There appears to be a higher degree of correlation between the story value of written compositions and comprehension than between the number of words written and the measures of comprehension. This is true even when the measures of comprehension are based upon reproductions and the reproductions are described in terms of the number of words or number of ideas reproduced.

7. In the measurement of rate of silent reading, the Courtis Silent Reading Test No. 2, is shown to have the highest degree of reliability. Monroe's Standardized Silent Reading Tests, which

were intended to yield only very crude measures of rate of silent reading, are shown to be among the most reliable tests.

8. In measuring comprehension, the Courtis Silent Reading Test, No. 2, is the most reliable.

9. The coefficient of reliability is shown not to be a satisfactory measure of reliability.

10. Comparisons with teachers' ratings indicate that, in the fourth grade, teachers tend to judge silent reading ability on the basis of the pupil's ability to answer questions. In the seventh grade, teachers give greater weight to the pupil's ability to reproduce or tell what they have read.

Correlation with composites. In Tables XVI and XVII, the corrected coefficients of correlation of each test with the composite scores are given. These, in general, are larger than the correlations between single tests. Monroe's Standardized Silent Reading Test correlates very highly with composite A. This means that this test, which is very simple to administer, yields measures of essentially the same traits as are secured by means of this composite, which in the fourth grade involves three scores and in the seventh, two scores. The correlation with composite C and with the general composite is also high. In fact, with the partial exception of Starch's Test, no other correlations are as high as these two composites of the Monroe Silent Reading Tests. It, therefore, appears, as judged by composite scores, that this test yields measures of comprehension which agree more closely with the composite measures secured from this group of tests than any other single test. The correlations for rate are also high.

BULLETIN NO. 9

BUREAU OF EDUCATIONAL RESEARCH
COLLEGE OF EDUCATION

WRITTEN EXAMINATIONS AND THEIR IMPROVEMENT

by

WALTER S. MONROE, Director

PRICE 50 CENTS

PUBLISHED BY THE UNIVERSITY OF ILLINOIS, URBANA
1922

UNIVERSITY
of ILLINOIS
PRESS

TABLE OF CONTENTS

PAGE
PREFACE 5
INTRODUCTION: Basis of a rational estimate of the value of written examinations 8

CHAPTER I. CRITICISMS OF WRITTEN EXAMINATIONS SET BY TEACHERS AND OTHER SCHOOL OFFICIALS 9
1. Examinations yield inaccurate measures of achievement 9
   A. Marking of examination papers subjective 9
   B. Questions of examination not equal in difficulty, and weighting by teachers subjective 11
   C. Content of examinations not in agreement with educational objectives 12
   D. Rate of work neglected 13
   E. Adequate opportunity for a pupil to demonstrate his ability not offered by single examination 13
   F. Marks assigned to examination papers imply subjective norms 14
2. Undesirable mental processes stimulated by examinations 16
3. Examinations tend to become educational objectives 17
4. Examinations injurious to health of students 17
5. Time devoted to marking of examination papers might be more profitably employed 17

CHAPTER II. THE DEFENSE OF WRITTEN EXAMINATIONS SET BY TEACHERS AND OTHER SCHOOL OFFICIALS 18
1. Measurement of abilities of students necessary to high degree of school efficiency 18
2. Substitutes for written examinations 18
   A. Standardized educational tests versus examinations 19
   B. Teachers' estimates versus examination marks 19
   C. Daily "grades" versus examination marks 21
3. Inaccuracy of examination marks 22
   A. Neglect of the rate of work not necessary 22
   B. Unequal difficulty of questions not a serious defect 22
   C. Inaccuracy of single examination 23

4. Examinations force students to review and organize content of course 24
5. Examinations furnish effective motive 25
6. Proper use of examinations 26
7. Examinations as objectives 26
8. Effect of examinations upon health of students 26
9. Time devoted to examinations a profitable investment 27

CHAPTER III. METHODS OF IMPROVING EXAMINATIONS 28
1. Increasing objectivity in marking of examination papers 28
   A. Increasing objectivity of "grades" by improving examination questions 28
      (1) General methods 28
      (2) New examinations: use of questions permitting only one correct answer 29
         a. True-false exercises 30
         b. "Yes" and "no" exercises 31
         c. Recognition exercises 31
         d. Completion exercises 34
         e. Other advantages of "new examination" 34
         f. Limitations of "new examination" 35
   B. Rules for marking examinations 36
2. Increasing objectivity of norms for translating examination scores into school marks 39
3. Securing closer agreement of examinations with educational objectives 41

CHAPTER IV. SUMMARY 42

APPENDIX 44
1. True-false examination in physiology 44
2. True-false examination in history and civil government 48
3. True-false examination in geography 52
4. Completion examination in American government 57
5. Recognition examination in algebra 59
6. Traditional examinations 62

PREFACE

During the past twenty years there have been many controversies concerning the value and place of written examinations. There have also been a number of investigations of examinations and examination "grades." Since standardized educational tests have become widely used a number of superintendents and teachers have proposed that they replace the written examinations set by teachers and other school officials. More recently some attention has been given to the improvement of written examinations by the application of certain principles of test construction.

Because of the importance of the written examination and also because a number of inquiries have been addressed to the Bureau of Educational Research, it has seemed wise to organize and publish a summary of the important ideas relating to both the criticism and the defense of examinations. To this there have been added a number of suggestions for the improvement of examinations. It is hoped that this bulletin may foster intelligent thinking relative to written examinations and their use in our schools.

Although this bulletin is largely the product of the labors of the Director of the Bureau of Educational Research it is only just that the contributions of other members of the staff should receive recognition. Both Mrs. Charles H. Johnston and Mr. Lloyd B. Souders have made substantial contributions.

Walter S. Monroe, Director

1

WRITTEN EXAMINATIONS AND THEIR IMPROVEMENT.

INTRODUCTION

Basis of a Rational Estimate of the Value of Written Examinations. Until recently, examinations occupied a regular place in the work of the school. Students expected them as a matter of course, and the accuracy of the marks placed upon examination papers was not seriously questioned. However, for a number of years written examinations set by teachers and by other school officials have been subjected to criticism. During this period the defects and the limitations of examinations have been thoroughly canvassed. Many prominent educators have advised that they be abolished entirely, and in a number of school systems this has actually been done. The friends of examinations, however, have urged their merits and have insisted that the abolition of them would cause our educational system to deteriorate. The controversy has not been without prejudice on both sides. The marking of examination papers involves much drudgery for instructors. Students dislike examinations partly because they require a type of intensive mental activity which many of them prefer to avoid and partly because it is fashionable in many schools to oppose them. Conservatives, naturally, have resented any proposal to change a system of education which they credited with producing the educated men of the present generation. Some, at least, have expressed the belief that examinations have been largely responsible for the quality of the output of our public schools and colleges.

In evaluating the criticisms and the defense of written examinations it is imperative that one keep in mind the fact that they have more than one function. Written examinations are not merely measuring instruments, although this function is probably most prominent in the thinking of many persons. The written examination is used as an instrument for measuring the achievements of students, but it also affords a unique type of opportunity for learning. Under rather well defined conditions, certain tasks are set for the pupil and he is required to demonstrate within a limited time what he is able to do. He is thrown upon his own resources and forced to work under pressure. In the actual writing of his answers to the questions of the examination the pupil has an opportunity to learn. Ideas tend to become more definite as a result of expression in written form. Frequently the pupil gains new ideas as a result of the reflective thinking he does in answering the questions. It is true that all pupils do not always learn in taking an examination, but it is also true that all pupils do not take advantage of all other educational opportunities which are offered them. In addition to the actual taking of the examination, the pupil frequently, as a preparation for it, engages in review; and, because he knows that later he must take the examination, he has a stronger motive for this review.

Not only is it important that we recognize the existence of functions other than the one of measurement, but it is also imperative that we bear in mind two distinctions. First, we must distinguish between criticisms of examinations and criticisms of certain kinds of examinations. The fact that some teachers set poor examinations does not furnish an adequate basis for concluding that all examinations should be abolished. In the second place, we should distinguish carefully between criticisms of examinations and criticisms of the ways in which they are used. Good examinations may be used for wrong purposes. For example, a good examination might be given to a pupil or a group of pupils merely as a punishment for some misbehavior. If we believe that such use is not justified it does not follow that the examination itself is subject to adverse criticism or that all examinations should be abolished.

CHAPTER I.

CRITICISMS OF WRITTEN EXAMINATIONS SET BY TEACHERS AND OTHER SCHOOL OFFICIALS

The arguments advanced for and against examinations have dealt with various phases. Some of the criticisms have emphasized the effectiveness of the examination as a measuring instrument; others have had to do with the purposes for which examinations are used by teachers and by other school officials. Some criticisms are based upon facts, while others merely represent opinions. In the following pages the most significant criticisms have been summarized and grouped under a few major heads. In presenting these criticisms there will be no attempt to point out their limitations or to present the arguments in favor of written examinations. This will be reserved until the second chapter.

1. Examinations yield inaccurate measures of achievement. A number of criticisms of written examinations set by teachers and by other school officials have referred to their effectiveness as instruments for measuring the achievements of students. These criticisms may be summarized under six heads.

A. Marking of examination papers subjective. Scientific investigation has proved that the marking of examination papers is subjective, i.e., different teachers, when working independently, tend to assign widely varying marks to the same paper. An investigation by Starch and Elliot[1] is typical of many that have been made. These investigators selected a final examination paper in geometry, written by a student in one of the largest high schools in Wisconsin. An exact reproduction of this paper and a set of the questions were sent to one hundred and eighty high schools in the North Central Association. It was requested that this paper be graded according to the practise and standards of the school by the principal teacher of mathematics. One hundred and sixteen acceptable replies were received. The papers showed evidence of having been marked with unusual care and attention. In seventy-three schools where the passing grade was 75 the lowest mark given was 39 and the highest 88. The mode was 75, with twelve teachers giving this mark. Of the one hundred and sixteen marks assigned to this paper, two were above 90 and one was below 30. Twenty were 80 or above and twenty other marks were below 60. Forty-seven teachers assigned a mark passing or above, but sixty-nine teachers thought this paper not worthy of a passing mark.

---

[1] Starch, Daniel, and Elliot, E. C. "Reliability of grading high school work in mathematics," School Review, 21: 254-59, 1913.

Robert L. Morton[2] reports an investigation of the reliability of the marking of examination papers written by teachers applying for a license to teach. In 1904, the Ohio Legislature provided for uniform questions for the teachers' examination. These questions were to be prepared in the office of the State Superintendent of Public Instruction and sent to the eighty-eight county boards of examiners. Special examiners were appointed in each county to rate the papers. Morton selected an arithmetic paper from the files of one board of examiners. The paper was mimeographed, care being taken to produce exactly the language, spelling, and punctuation of the original paper. A copy of this paper, together with the questions, was sent to each of the eighty-eight county superintendents in Ohio with the request that it be graded by the special examiner for arithmetic. Replies were received from fifty-five counties. The lowest mark given to the paper was 60 and the highest 99. In marking the answer given to one question on this paper five examiners rated it at zero, twenty-one at 10, and the other twenty-nine assigned marks between these extremes. If each answer had been rated in the county assigning the lowest mark to it, the total "grade" for the paper would have been 28. On the other hand, if the highest marks assigned to the answers of the various questions had been used to make up a "grade", a mark of 100 would have been given to the paper. Morton investigated in a similar way the marking of a paper in the theory and practise of teaching and also of one in geography. Similar variations in the marks were found.

A striking illustration of the subjectivity of the marking of examination papers by college instructors is cited by a recent writer[3]. One of the group of expert readers assigned to the marking of examination papers in history, after scoring a few papers, wrote out for his own convenience what he considered model answers to the questions. By some mischance this "model" examination paper fell into the hands of another expert reader who graded it as a paper written by a student. The mark he assigned to it was below passing and, in accordance with the custom, this "model" was rated by a number of other expert readers in order to insure that it was properly marked. The marks assigned to it by these readers varied from 40 to 90.

---

[2] Morton, Robert L. "The examination method of licensing teachers," Educational Administration and Supervision, 6: 421, November, 1920.

[3] Wood, Ben D. "Measurement of college work," Educational Administration and Supervision, 7: 301-34, September, 1921.

Scientific investigation of the marking of examination papers has been sufficiently extensive to prove that, except in a very few instances, the process is subjective. Except for accidental errors, different teachers should assign the same mark to an examination paper in spelling. The marking should also be highly objective in arithmetic unless there is an attempt to allow partial credit for examples and problems partially right or for correct principle when the answer is not correct. The marking of the answers to questions which call for specific facts, such as dates, names of places, or persons, should approach objectivity. With the exception of these cases, the marking as it is ordinarily done is highly subjective, and hence the "grades" are inaccurate measures of achievement.

As might be expected, the degree of subjectivity varies with different school subjects. It is, however, sometimes found to be high where the nature of the subject matter leads one to expect that the marking will be relatively objective. For example, Starch and Elliot found that the marking of an examination paper in geometry was just as subjective as one in English or history. Kelly[4] found that the rating of examination papers in algebra was considerably more objective than in physics.

B. Questions of an examination not equal in difficulty, and weighting by teachers subjective. There is abundant evidence that the questions of an examination are generally not equal in difficulty. Frequently, in this respect, they vary widely. When the questions are submitted to a large number of pupils, some will be answered correctly by a large percent of the pupils, others by only a small percent. To give as much credit for answering an easy question as for a difficult one would appear to introduce serious errors into the marks assigned to the papers. Because it is recognized that the questions which make up an examination are generally unequal in difficulty, teachers frequently attempt to assign appropriate weights. For example, one question may be assigned a credit of 15 points while an easy one is given a credit of only 4 points. One investigation[5] has shown that teachers' estimates of the difficulty of questions are highly subjective. Twenty teachers were asked to arrange twenty-three problems in arithmetic in the order of their difficulty. A very wide variation in these rankings was found. One problem was considered the easiest by one teacher and ranked twenty-first in difficulty by another. The results of this investigation seem representative. That being the case, any weighting of questions by teachers must be considered highly subjective, and hence not a satisfactory corrective for the unequal difficulty of questions.

---

[4] Kelly, F. J. "Teachers' marking," Teachers College, Columbia University, Contributions to Education, No. 66, 1914.

[5] Comin, Robert. "Teachers' estimates of the ability of pupils," School and Society, 3: 67-70, January 8, 1916.

C. Content of examinations not in agreement with educational objectives. The criticism is frequently made that teachers, in formulating examination questions, tend to ask for unimportant details and to neglect the minimum essentials of a subject, and that, therefore, a pupil's performance on an examination can not be a truthful index of the extent to which he has achieved the educational objectives set for him. Some questions are described as "catch questions." By this, it is usually meant that such questions call for some unimportant detail or that they are ambiguous in some way. There appears to have been no scientific investigation of the character of the examination questions asked of pupils. However, it is doubtless true that this criticism has justification in some cases because frequently teachers give relatively little time to the preparation of their questions, and these often reflect any hobbies or prejudices which the teachers may have. Experience in the construction of standardized educational tests has shown that it is difficult to eliminate all ambiguity and indefiniteness in questions. Hence, it is likely true that many questions are not well stated, and for this reason are not properly understood by those taking the examination. When this is the case, the "grades" tend to be inaccurate measures of achievement.

When an examination is set by some person other than the teacher of the class it not infrequently happens that many of the questions pertain to topics which have received little or no attention during the instruction periods. In many schools it seems to be the custom for the superintendent or the principal, without consultation with the teacher in charge, to make out the questions for the final examination on which the pupils' semester grades are largely based. For example, in a fifth grade geography class in an Illinois city, four of the five questions of the examination concerned current conditions about which the children, instructed only in their texts, knew little. A few pupils, fortunate enough to have heard these matters discussed in their own homes, received a passing grade. The majority of the class failed. This examination, interesting and in itself not subject to criticism, should not have been used, however, as a means for measuring the achievements of that particular class. It was not in agreement with the educational objectives toward which the teacher had directed their efforts. Such examinations are "hard" in the sense that capable students will answer only a relatively small percent of the questions correctly, and are rightly criticized as being unjust because the students are not given an opportunity to demonstrate their achievements.

D. Rate of work neglected. The usual plan is to set an examination which practically all pupils can finish in the time allowed. No record is kept of the time which the pupil has spent in writing his answers. If two pupils write papers which are considered equivalent in quality but one has completed the examination in forty minutes and the other in ninety minutes, it is not customary to distinguish between their performances. Both will receive the same "grade." This means that the rate of work is neglected. Since the rate at which a pupil is able to answer questions is one index of his ability, the ordinary examination fails in this respect to secure a truthful measure of his ability.

E. A single examination does not offer an adequate opportunity for a pupil to demonstrate his ability. Some critics urge that a single examination, even when carefully prepared and graded, will not in general yield a reliable measure of a student's ability. McAndrew[6], in reviewing the work of the New York City high schools, says, "New York City high schools use them (examinations) in deciding the promotion to the training school for teachers. We have every year some students whom their teachers have complimented regularly but who fail of graduation because of a three hour test which nullifies the work of four years. I cannot see how a pupil writing for three hours can be tested for what he has done for a year or more." Courtis[7] expresses much the same thought in the following statement: "The best examination is not that represented by the score of a single performance in a single day. Human effort is variable and human skill too easily upset to make it fair to have promotion based upon chance scores."

Thorndike[8] has summarized a number of investigations carried on at Columbia University in order to determine the reliability of the "grades" made on college entrance examinations as a basis for predicting the type of work which the student will do in college. He states that we cannot estimate the success of the student in college from his grades on entrance examinations with "enough accuracy to make the entrance examinations worth while and to prevent gross injustice being done to any individual. The record of eleven or more entrance examinations gives a less accurate prophecy of what a student will do in the latter half of his college course than does his high school record." Similar results have been obtained by other investigators.[9] Studies of this type do not necessarily prove that examination "grades" are inaccurate measures of achievement. Because of other elements which enter into college life, a student having made a satisfactory record in his secondary school may not carry on successfully his work in college.

---

[6] McAndrew, Wm. "Our old friend the examination," Proceedings of National Educational Association, 1916, pp. 527-33.

[7] Courtis, S. A. "Standardizing of teachers' examinations," Proceedings of National Educational Association, 1916, pp. 1078-86.

[8] Thorndike, E. L. "The future of the college entrance examination board," Educational Review, 31: 470-83, May, 1906.

F. Marks assigned to examination papers imply subjective norms. This criticism has to do with errors in interpreting measures of achievement yielded by examinations rather than with errors in the measures, but since examination marks are usually thought of as measures, it may be considered under this head. In reading the controversial literature on written examinations one will find little mention of the criticism expressed by the heading of this paragraph, but in the illustrations given below the reader will recognize that this weakness of examination marks has been sensed by most pupils and teachers. However, they appear to have failed to analyze the situation sufficiently to grasp the source of the difficulty. In the judgment of the writer this is one of the most serious weaknesses of the traditional examination.

In order to understand how norms (standards) are used in connection with the grading of examination papers it is necessary to distinguish between scores, or measures, and "grades," or marks. A score simply describes the performance which has been recorded in the examination paper. For example, a pupil may answer 55 per cent of the questions correctly. In this case 55 is his score. If a certain number of points or credits had been given for each question his score might be 129 or 91, or 217. A "grade" interprets this description with reference to certain norms. A "grade" indirectly describes a pupil's performance on an examination, but it also tells whether the pupil's performance is to be considered as above passing or below passing; whether he is to receive the highest mark or the lowest mark or an average mark. It is customary to describe the quality of examination papers in terms of the percent of questions answered correctly. For example, if an examination includes ten questions and a pupil answers seven of them correctly and an eighth one partially right, he is given a score of 75 per cent, which is interpreted to mean that in the judgment of the examiner he has answered the questions 75 percent correctly. School marks or "grades" are also frequently expressed in terms of percents. Sometimes they are expressed in terms of letters or other symbols, but these in turn are defined in terms of percents. For example, the grade of "A" may be defined as being between 95 percent and 100 percent.

---

[9] Lincoln, E. A. "Relative standing of pupils in high school and early college, and their college entrance examinations," School and Society, 5: 417-20, April 7, 1917.

Since both scores and "grades" are generally expressed in terms of percents, it is only natural that the two have been confused and that scores have been used as "grades." A good illustration of their difference came to the writer recently. An examination in mathematics was given to nearly 1000 freshmen in one of our large universities. This examination may properly be described as "hard," considering the training which the students had received. One student made a score of 100. The lowest score was 12. The average was approximately 55. From the standpoint of the distribution of scores this was a "good examination." If it had been easier, so that any considerable number of pupils received scores of 100 percent, it would have been defective. If it had been so "hard" that a considerable number of students made zero scores it would also have been defective. In both cases it would have failed to differentiate between some students who were not equal in ability. However, it is obvious that if a passing mark of 70 or 75 is adopted, it would be unjust to say that all pupils having scores below this passing mark should receive a "grade" of failure. The passing mark for this particular examination should be in the neighborhood of 40. If in this case it is desirable to have the scores represented in terms of "grades" a score of 40 should be translated into a "grade" of 70 or whatever passing mark this institution has adopted.

The recognition of this distinction between scores and "grades" enables us to indicate the way in which subjective norms are implied in "grades". A "grade" is not a pure measure or description of the pupil's performance. It is rather an interpretation of the measure of his performance with reference to certain norms. When no distinction is made and scores are used as "grades," pupils will receive high "grades" if the examination is "easy;" if it is "hard" they will receive low ones. Thus, the difficulty of the examination is one factor in establishing the norms with reference to which the scores are interpreted when they are used as "grades". Severe marking will tend to set high norms. It is only when the examination is of average or "standard" difficulty and the marking is average in severity that scores and "grades" become identical in magnitude. Since the norms are established by the difficulty of the examination and the severity of the scoring, they must be subjective. In the investigations of the marking of examination papers it was shown that teachers varied widely in their judgments concerning the worth of examination papers. There is no reason to expect that they would agree more closely in estimating the difficulty of examinations. Hence, norms which depend upon teachers' estimates of which questions are appropriate for examinations and upon their marking of the papers must be considered subjective.

2. **Undesirable mental processes stimulated by examinations.** A number of critics have urged that examinations—and especially the preparation for them—tend to stimulate undesirable mental processes, and that as a result a student's learning is not of the right kind. It is claimed that memory is emphasized to the exclusion of higher mental processes. Many questions call only for facts, and even in the case of those which require reasoning the student must have some facts with which to reason. He is generally asked to answer the questions without having access to his text books or to other sources of information. It is seldom that the examination provides the student with any of the facts or principles which he requires in the reasoning process. Hence, what he records upon his examination paper must be based upon his own store of information. Thus, it is undoubtedly true that examinations make heavy demands upon the memory of the pupils.

Those who have supported this criticism of examinations insist that this emphasis upon memory tends to change the child's mind into an automatic machine. The following statements are typical: "It is mere commonplace to say that in many schools the best preparation for an examination is to have in memory a vast number of details. The student may undertake the examination in such a condition of brain fatigue that he would find it difficult to solve a simple original problem; but if he has this plethoric memory of details he will succeed."[10] "The memory which gives back knowledge in the exact form in which it is received is likely to shine at examinations."[11]

In this connection it is claimed examinations place a premium upon "cramming." Since memory is emphasized it is only natural that students should engage in an intensive review of the course immediately preceding the examination period. It is a well established principle that the recency of an experience materially affects its recall. The critics of examinations insist that students who have given little attention to the work of the course during the term may write a creditable examination paper merely as a result of an intensive cramming immediately preceding the examination date.[12] Since this is true, students tend to loaf during the term and then "cram" for examinations.

When a student's learning is limited to a brief period of intensive cramming immediately preceding the examination, he will tend to forget very rapidly. In addition, he has missed the regular growth that was possible during the term and has had no opportunity to organize and relate the various items of the course.

---

[10] Monroe, Paul. Encyclopedia of Education, Vol. 2, 536-38.

[11] Schaeffer, N. C. "The unmeasurable in teaching," N. E. A. Proceedings, 51: 169-72, 1913.

[12] Hollister, H. A. High School Administration. Boston: D. C. Heath & Company, 1909.

3. **Examinations tend to become educational objectives.** It has been charged that teachers tend to emphasize examinations in such a way that the students come to think of them as the educational objectives to be attained. Because the fear of making a low "grade" on an examination is an effective motive, teachers frequently use it. The students are reminded of approaching examinations and are warned that they will not be able to pass them unless they study certain topics. When the "grade" received upon the final examination determines the student's success in the course, it naturally follows that he will think of "passing the examination" as the objective to be attained. Sometimes students secure former examination questions or compile a list of the topics the instructor has emphasized and study these to the exclusion of other topics which may be more important. Thus, the real objectives to be attained are obscured.

4. **Examinations injurious to health of students.** Because of the strenuous preparation for examinations and also because of the obvious strain which accompanies the taking of an examination, a number of critics have urged that examinations are injurious to the health of students. Even students who have been faithful during the term will frequently cram for the examination because of a desire to take a high rank in their class or because of the additional credit for quality that is given in some institutions. When a student is required to write for a number of hours during a single day, as sometimes occurs, it is obvious that even under the best conditions he has expended a large amount of nervous energy.

5. **Time devoted to marking of examination papers might be more profitably employed.** A majority of our larger colleges and universities have recognized the fact that the marking of ordinary examination papers makes heavy demands upon the time of instructors and have provided "assistants" or "readers" who are to rate the papers. Not infrequently the examination period is dreaded by instructors as well as by students because of the drudgery involved in reading examination papers. One writer[13] has estimated that it requires approximatly three hours to construct, give, and mark an examination for an average class. A teacher will give twenty or more examinations per year. This means that each teacher devotes not less than sixty hours to the construction, giving, and marking of examinations. This estimate is conservative. Many teachers spend much more time than this. The critics of examinations have pointed out that this time might be more profitably devoted to other school activities.

---

[13] McCall, W. A. "New kind of school examination," Journal of Educational Research, 1: 33-46, January, 1920.

CHAPTER II.

THE DEFENSE OF WRITTEN EXAMINATIONS SET BY TEACHERS AND OTHER SCHOOL OFFICIALS

The criticisms of written examinations summarized in the preceding chapter appear to make a strong case against their use in schools. Some of the criticisms are established facts, and others corroborate our general observations. It is, however, necessary to remember that only one side of the case has been presented. Some of the criticisms are much less serious than their advocates would have us believe. Others are not fundamentally criticisms of examinations, but rather of the way in which they are used and of other phases of our school procedure. In addition, certain significant merits of examinations were not mentioned. In this chapter the criticisms set forth in Chapter I will be examined, and the defense of examinations presented. In Chapter III we shall explain certain methods of improving written examinations which tend to eliminate or greatly minimize certain defects that now exist.

1. **Measurement of abilities of students necessary to high degree of school efficiency.** In considering written examinations as measuring instruments, it is necessary to bear in mind that the work of the school cannot be carried on in an efficient manner without from time to time measuring the abilities of students. These measurements are essential to the organization and administration of our schools as well as to the instruction of the class room. The promotion and classification of pupils, the guidance of pupils, both educational and vocational, the supervision of instruction, and even instruction itself, cannot be most efficient unless the abilities of the pupils are measured at intervals. If these measurements are not made in one way they must be in another. If written examinations are abolished the abilities of pupils must be measured in other ways, or the efficiency of our schools will decrease.

2. **Substitutes for written examinations.** It has been proposed by some critics that substitutes for written examinations may be used. Standardized educational tests, teachers' estimates, and daily "grades" furnish types of information upon which the measurement of the pupil's achievement may be based. However, it would seem, for the following reasons, that such means of measuring cannot be considered as satisfactory substitutes for written examinations.

A. Standardized educational tests versus examinations. Standardized educational tests have been shown to be superior to ordinary examinations as measuring instruments, but in considering the proposal that they replace written examinations certain facts must not be overlooked. The number of satisfactory standardized educational tests is as yet very limited. In only a few school subjects, such as handwriting, spelling, arithmetic, and oral and silent reading, do we have available standardized educational tests which might be used as substitutes for written examinations set by the teacher. Even in these subjects it may happen that for some reason the teacher has emphasized certain topics and omitted or treated casually others. Furthermore, official courses of study vary widely from city to city. Standardized educational tests are of necessity confined to those topics which are uniformly taught or, at most, to those topics which are generally taught; therefore, in a particular instance there may not be available any standardized educational test which is adapted to the instruction which the class has been receiving. A teacher, however, can construct an examination which is specifically designed to measure the results of instruction given to a particular class. Thus, standardized educational tests can be used only to a limited extent as substitutes for examinations set by the teacher or by other school officials.

In this connection it is well to remember that standardized educational tests are not perfect measuring instruments. Even the best of them do not begin to approach the degree of accuracy to which we are accustomed in the measurement of physical objects. The results which they yield may involve errors so large as to distort or make entirely erroneous many of our interpretations. Thus, in considering the errors which are involved in examination "grades" we should not forget that our best instruments for measuring mental abilities are far from perfect with reference to accuracy.

B. Teachers' estimates versus examination marks. Some of the critics of examinations have insisted that experienced teachers are able to estimate the achievements of students by reason of their acquaintance with them during the term. It has been claimed that these estimates will be more truthful measures of achievement than the "grades" obtained from a written examination given at the end of the term. There is no doubt that experienced teachers can, under favorable circumstances, estimate with considerable accuracy the achievements of students. If the class is reasonably small and if the teacher has used methods of instruction which have called for frequent oral and written performances by the students and has kept a careful record of the quality of these performances throughout the term, the estimates may be relatively accurate measures of the achievements of the students. However, there are certain limitations which should be noted.

Teachers may be unduly influenced in their estimates by the recent performances of the students. Unless careful records have been kept throughout the term, inferior performances at the beginning tend to be overshadowed and the teacher's final estimate based upon the work of the last few weeks. In case the class is a large one, the teacher does not have an adequate opportunity of becoming intimately acquainted with the students. Teachers' estimates are likely to be materially affected by personal characteristics of pupils. A pupil with a pleasing personality or one who is liked by the teacher is likely to be rated higher than one who is unattractive.

If the class work is conducted so that there is little or no written performance on the part of the students, teachers' estimates will necessarily be based almost wholly upon the oral responses which students give during the class period. Some pupils make a good showing in class when the recitation is oral, but are at a decided disadvantage when asked to record their answers in writing. Frequently this difficulty is due to the fact that they are careless in their thinking and do not have well formed ideas to express. In oral recitation they are able to make a fair showing because of personal characteristics and because of the stimulus of detailed questioning by the instructor. Furthermore, in a class discussion a bright student who has a good command of language may easily pick up ideas from other members of the class and recall items from his general experience sufficient to make a good showing. On the other hand, there are students who express themselves most effectively in writing. They may be good thinkers but a little slow in their mental processes and not clever in discussion. Thus, it is difficult or impossible for teachers to estimate accurately the real achievements of students from oral recitations alone. In classes where a large amount of written work is required, as in English composition, the teacher has an opportunity, in formulating estimates, to consider the pupil's written as well as oral performances. In such subjects, the proposal to replace written examinations by teachers' estimates based upon the work of the course has much more merit than in other subjects where there has been only a small amount of written work. But even when much daily written work is required, a written examination covering the work of the term may give additional information concerning the real achievements of some students.

C. Daily "grades" versus examination marks. It has been proposed that the best measure of a student's achievements throughout a term is furnished by the average of his daily "grades." These "grades" include teachers' estimates of pupils' performances in class, marks given for written work submitted, and "grades" earned on short quizzes. Those who favor using the average of these marks as a pupil's final grade point out that by this means we secure a measure of a pupil's achievement which depends upon the quality of his work day by day and which, furthermore, is the average of a large number of measurements. Probably no one would contend that a student's daily "grades" should not receive consideration in determining his final standing. There are, however, certain dangers in taking the average of a pupil's daily "grades" as his final mark. In the first place, if the student knows that his final mark depends only upon his daily work he is likely to study for the day alone. His daily "grade" is based upon a rather small unit of work. Furthermore, it is based upon his performance immediately after studying the assignment. Thus, a pupil's ability to organize the content of a course and his retention of it are likely to receive little consideration in making up his final mark. Of course, it is possible for a teacher to ask the student to review frequently and to summarize and organize the work of the course at the end of the term. However, it is difficult to convince students of the necessity of reviewing the work of a course if the performance on this review receives no more weight in determining the final "grades" than the performances during an equal period of time elsewhere in the course. When classes are large it is difficult to have each student recite each day, at least to a sufficient extent to furnish any basis for assigning a daily "grade." When the members of the class are called upon only occasionally it frequently happens that they prepare only those lessons upon which they think they will be questioned. Of course, a resourceful instructor can do much to discourage this practise but the fact remains that daily "grades" may not be representative samples of a student's achievements in a course. A written examination covering the work of the entire course or of a large division of it will furnish a measure of achievement which can be secured in no other way.

Conclusion: No satisfactory substitute for written examinations. Teachers' estimates and daily grades furnish certain types of measures of the achievements of pupils. In certain cases standardized educational tests can be used to secure relatively accurate measures of achievement in certain fields. However, these means of measuring cannot be considered satisfactory substitutes for written examinations. The written examination yields a type of achievement which cannot be secured through any other means. This measurement is more important for some school subjects than for others but in relatively few is one justified in abolishing the written examination as a means of measurement.

3. **Inaccuracy of examination marks.** In Chapter I several sources of error in examination marks were mentioned. In the case of some of these sources sufficient experimental evidence was presented to be convincing. The only way in which these sources may be eliminated or minimized is by modifying examinations. These modifications will be treated in Chapter III, but three sources of error may be commented on at this place.

A. **Neglect of the rate of work not necessary.** It is not necessary that the rate of work be neglected in examinations set by teachers and other school officials. It is easily possible to take into account a student's rate of work in determining the mark which he is to receive on his examination paper. One plan is to set an examination of sufficient length to keep all members of the class employed during the entire period. In marking the papers, the number of questions which the student has answered should be considered as well as the quality of his answers. Another procedure is to have each student record the time when he finishes. This record will enable the teacher to determine the amount of time devoted to answering the questions. In the measurement of handwriting and of silent reading, as well as of abilities in a number of other fields, it is extremely important that the rate of work be considered.

B. **Unequal difficulty of questions not a serious defect.** It does not appear that accurate measurements of the abilities of students can be secured by giving the same credit for answering an easy question as for answering a difficult one. However, investigations of this question in connection with the scoring of standardized educational tests have indicated that the errors introduced by this procedure, which appears to be illogical, are not large. After having weighted the exercises of his language and grammar tests on the basis of difficulty, Charters[1] dropped the weights because he found that the correlation between the weighted and unweighted scores was slightly over 0.90. A number of other test makers have likewise used exercises which were unequal in difficulty without assigning any weighted credits to them. A number of other tests which consist of exercises arranged in ascending order of difficulty have been scored by taking the number of exercises done correctly, which amounts to giving as much credit for doing an easy exercise as for doing a more difficult one.

---

[1] Charters, W. W. "Constructing a language and grammar scale,"Journal of Educational Research, 1:249–58, April, 1920.

The writer[2] has shown in the case of certain standardized educational tests that the error introduced by disregarding the inequality of the questions in scoring is not larger than the variable error of measurement due to the variability of the performances of pupils. It, therefore, seems likely that the error introduced by disregarding the unequal difficulty of the questions of an examination will not be significant in most cases. When the examination consists of as many as twenty or more questions the unequal difficulty of the questions probably can be safely disregarded. Of course, it is wise to avoid extreme variations in difficulty whenever possible. Those setting examinations should endeavor to formulate questions which do not differ widely in this respect.

There is some doubt whether the difficulty of a question can furnish a criterion for determining the amount of credit which should be given for answering it correctly. Achievements are not valuable merely because they are difficult but because of the relation which they sustain to our educational objectives. Questions which are relatively unimportant may be very difficult. In fact, lack of instruction is one reason why a question is difficult. Therefore, topics which are assigned minor places in a course of study may furnish the basis for very difficult questions, while the topics which have been emphasized because of their importance may furnish the easier questions. It is not unlikely that the best basis for weighting questions is their social importance rather than their difficulty.

C. Inaccuracy of single examination. It has been urged that a single examination cannot yield an accurate measure of the achievements of students. It is human nature to be variable in one's performances. For certain reasons students whose achievements are ordinarily high may make low "grades" on a single examination. It must therefore be admitted that many of the "grades" yielded by a single examination are likely to involve serious errors. However, in considering the significance of these errors with reference to the use of written examinations, it is necessary to distinguish between examination "grades" and final "grades." The fact that the "grade" for a single examination may be inaccurate and, hence, that a final "grade" should not be determined by a single final examination is not sufficient reason for abolishing written examinations. It simply means that more than one written examination should be given or that other measures of achievement should be taken into consideration in determining final "grades." Teachers' estimates and daily grades should have a place. Any written tests which may be given during the term should also be given some weight. A final examination, however, deserves a place also. In some subjects it should be given more weight than in others.

---

[2] Monroe, Walter S. "A discussion of the performances of pupils on exercises of varying difficulty," School and Society, 15:341–43, March, 1922.

4. **Examinations force students to review and organize content of course.** One of the criticisms made in Chapter I was that written examinations tended to emphasize memory and consequently to minimize the opportunities for the functioning of the higher mental processes. On the other hand, the friends of written examinations insist that they tend to produce exactly the opposite effect. They contend that final examinations have a positive value in that they tend to force students to review and to organize the content of a course. This is a very important phase of learning in many subjects. It is pointed out that the emphasis which examinations place upon memory is not undesirable. Students should learn many facts, and they should learn them with sufficient accuracy and in sufficient detail to answer most examination questions. The memorization of a considerable number of facts is to be commended. Even the intensive cramming which is deplored by many persons has educational value. If a student studies throughout the course, an intensive review and organization of the material is extremely valuable. In fact, the profit to the student is probably relatively greater for time expended in this work than for any other equal amount of time devoted to the subject. Unless carried to an extreme, cramming is undesirable only when it has not been preceded by thoughtful study. Even in this case it is better to have the student "cram" for the examination than to go through the course without engaging in any learning. Incidentally, it should be noted that the existence of "cramming" is not the fault of the examinations but of the type of instruction given and of the policy of making the student's final grade depend wholly or very largely upon his examination mark. With proper instruction and with considerable weight given to his record throughout the course, it will not be possible for a student to loaf during the term and secure a creditable final "grade" by "cramming" immediately preceding the examination. The fact that some students are able to earn high marks merely by a brief period of intensive study immediately preceding the examination is more of a reflection upon the content of the course than upon examinations. Undoubtedly, there are some students who are sufficiently brilliant to learn within a period of a few days as much

as the average student learns throughout the course; and, furthermore, it is probably true that in some of our courses relatively little is demanded of students.

It may be pointed out, also, that the extreme emphasis upon memory cited by some critics is not the fault of examinations in general but of particular types of questions. It is possible to ask questions which may be answered entirely from memory. It is also possible to ask questions which require much reasoning and organization of information.

The writing of an examination may itself be an important part of the student's learning. "There is no impression without expression." The writing of a three hour examination is undoubtedly an intensive form of expression. In case the questions are such that the reasoning and organization of information are required, the student may learn a great deal from the act of writing his answers to the questions. From this point of view, it may be urged that a student who is not required to take final examinations is deprived of an important opportunity for learning. Not infrequently students who have been "excused from examinations" in high school report that they experience a distinct handicap when they become college students.

5. Examinations furnish effective motive. Although the written examination should not be defended merely for the reason that it furnishes an effective motive, it must be admitted that this motive makes a special appeal to many pupils. In so far as possible, the pupils should be motivated by the intrinsic values of the subject matter. However, these values are frequently remote from the life of the student, and it is necessary to resort to motives which make a more immediate appeal. This point of view is effectively set forth in the following quotation: "Something in addition to merely appealing to the future motives seems necessary at times to urge along the average reader, some means whereby his future means are made comparatively more immediate. Not merely does the examination stimulate the reader generally, but it also helps to determine his specific interest and behavior. If the child feels himself compelled to give an account of what comes within the field of his perception, he will soon develop a habit of attending to that for which he will probably be held accountable. The examination, when properly applied, constantly reminds the reader that he is responsible for the mastery of what is proposed for his learning."[3]

---

[3] Myers, G. C. "Examinations and the learner," Educational Review, 54:274-84, June, 1917.

6. Proper use of examinations. As we have just pointed out, examinations may fulfill a very important function in the education of students. Also they may be abused. In some schools final examinations are used as a punishment. Students who attain a fixed average in their daily grades and who have not fallen below a minimum in their deportment are excused from final examinations. In such a case the examination is regarded both by teachers and students as undesirable or evil and naturally fails to fulfill its function. This practise of excusing some students from examinations is not a defensible one. The implication is that the examination is not an integral part of the school work, but merely a punishment or a last resort measuring instrument. It is probably wiser to abolish final examinations than to follow the practise of excusing students upon the basis of high daily grades and satisfactory conduct. In considering the defense of the written examination it is necessary to bear in mind that the existence of such abuses of examinations should not be interpreted as criticisms of examinations, but rather of the teachers and other school officials who use them.

7. Examinations as objectives. Examinations have been criticised because they tend to become immediate educational objectives. This is not wholly undesirable. If the examination consists of appropriate questions a pupil who studies to pass the examination will be directing his energies to appropriate ends. It is only when the examinations tend to become exclusive objectives or consist of inappropriate questions that this criticism has much force. Furthermore, it should be recognized that this criticism also is not a criticism of examinations as such, but rather of the use that is made of them by teachers and other school officials.

8. Effect of examinations upon health of students. While it is undoubtedly true that the preparation for examinations and the writing of them frequently make enormous drains on the energies of students, we have no carefully conducted investigation of the actual effect upon their health. To one who observes the enormous expenditures of time and energy which students devote to social and athletic activities, it is difficult to believe that examinations are in general more injurious to the health of students than many other activities in which they are permitted and even encouraged to engage. It is undoubtedly true that in the case of some individual students examinations make a heavier drain upon their energies than they should be asked to bear. However, here again it should be realized that this criticism is not fundamentally the criticism of examinations but rather of setting very long examinations or of placing extreme emphasis upon them by making the final "grade" depend wholly or very largely upon the examination "grade."

9. Time devoted to examinations a profitable investment. After a teacher has instructed his students to the best of his ability some will have failed to learn; some will have learned certain details but neglected others; and a few may have mastered all of the assignment. It becomes necessary at some time for each teacher to check up the work of his students, in order to know which students need supplementary or remedial instruction. Only by thus ascertaining the extent to which his students have achieved can the teacher make his instruction fit the needs of his class. For these reasons the time which is devoted to the construction of an examination, giving it, and rating the papers should be considered a legitimate expenditure of the teacher's time; in fact, in some cases it is doubtless among the most profitable of the time expenditures which the instructor makes.[4] In fact, the instructor may very profitably do more than simply rate the examination papers. In the case of students receiving low marks, their answers may be studied in order to ascertain in what ways and why they have failed. Such information will frequently be exceedingly illuminating to the instructor. This argument applies more forcibly to examinations or "tests" given at intervals throughout the course than to final examinations, but even in the latter case the instructor will frequently receive many helpful suggestions for improving his teaching.

---

[4] This argument applies only to a very limited extent to examinations given at the end of a course.

CHAPTER III

METHODS OF IMPROVING EXAMINATIONS

1. Increasing objectivity in marking of examination papers. The principal source of error in the "grades" assigned to examination papers is the subjectivity of the marking. This fault has been sufficiently demonstrated. It is, however, possible to increase materially the objectivity of written examinations. This may be accomplished in two ways: (1) improving the examination questions, and (2) introducing improved methods of marking examination papers.

A. Increasing the accuracy of "grades" by improving examination questions.

(1.) General Methods. The questions should be so stated that all pupils will interpret them in the same way. When a question is ambiguous the performances which different pupils give are not comparable. Pupils who are so unfortunate as to interpret a question in a way not intended by the instructor, can receive no credit for their answers unless the instructor recognizes the various interpretations. Even in this case all answers are not comparable because some are judged with reference to one interpretation and others with reference to another. A striking illustration of the difficulty of eliminating all ambiguous questions occurs in Form 7 of the Army Alpha Intelligence Test. The exercise asks the person taking the test to tell whether "cleave" and "split" mean the same or opposite. Both answers are correct because "cleave" has two meanings which are exactly opposite. Apparently this ambiguity was not detected although the test was prepared with care by well trained persons. However, in most cases, teachers will be able to avoid ambiguous exercises if they are subjected to a careful scrutiny. All questions should be so stated that they will be clearly understood.

Pupils should be given definite instructions concerning the methods of work to be employed. They should be told whether they are to work rapidly or slowly. In some cases it is advisable for the examiner to say at appropriate intervals, "It is now time to begin on question ——." Perhaps students should not be forced to comply with these directions, but they will tend to prevent an improper distribution of time over the list of questions. In questions in which pupils are asked to "discuss" or "explain" the completeness of the discussion desired should be indicated. In arithmetic pupils should be given directions concerning the form in which their work is to appear. If it is to be copied in a particular form all should do this; if it may be left in the form used to calculate the answer all pupils should leave it in this form.

From the standpoint of securing accurate measurements it is best to set an examination which is relatively hard. If more than 10 percent of the class make perfect "grades" the examination is "too easy," i.e., it is not difficult enough to yield accurate measures of the achievements of bright pupils. Unless the class is a selected group it is not possible for more than 10 per cent of the best pupils to possess exactly the same degree of achievement. The examination should be long enough so that all pupils will be employed for practically the entire period. The examination is not only useful as a device for measuring the improvements of pupils, but also valuable as a school exercise. No pupil should be permitted to waste any of the time set aside for the examination.

(2.) **New examinations: use of questions permitting only one correct answer.** The marking of examination papers becomes subjective when the scorer is asked to exercise judgment in determining the credit to be given for the pupil's performance. In spelling, a pupil's performance is either right or wrong and our practise is to allow no credit for a performance which is not entirely correct. Thus, the marking of an examination paper in spelling is highly objective because the scorer is not asked to exercise judgment concerning the quality of the performance. A high degree of objectivity may be attained also in the operations of arithmetic by agreeing to give no credit for examples partly correct. In other subject matter fields we are accustomed to ask some questions which call for specific facts and, hence, admit of only one correct answer. It has been claimed that such questions appeal only to the pupil's memory and that they do not yield an index of his acquaintance with principles and of his ability to organize and apply his knowledge. In order to reach this phase of his education we have asked the pupil to "discuss," "tell why," "compare," etc. When a pupil is asked to formulate an answer consisting of one or more sentences it is difficult or impossible to classify the performance as either right or wrong. When scorers are asked to exercise judgment in evaluating such performances, wide differences of opinion exist. In order to overcome this subjectivity of marking, it has recently been proposed that we can measure a pupil's acquaintance with principles and ideas by means of certain types of exercises which permit of only one correct answer. Such exercises have been used in our standardized educational tests, and it is now suggested that they be used by teachers in their examinations. Four types of such exercises will be considered.

a. **True-false exercises.** Instead of asking the pupil to formulate an answer in response to a question, we ask him to tell whether a given statement is true or false. For example, instead of the question, "Why did the Puritans come to America in the 17th Century?" we ask whether the following statement is true or false, "The Puritans came to America in the 17th century seeking wealth." The pupils may give his answer to this true-false exercise by writing a plus sign if he considers it true and a minus sign if he considers it false. The mental processes required in answering such exercises do not appear to be the same as those which occur in answering questions of the usual type. However, experimental evidence indicates that there is a high correlation between the scores which pupils make on a true-false examination and their acquaintance with ideas and principles as determined by our ordinary examinations.

Directions for constructing a true-false examination.[1] 1. In constructing true-false exercises, one may prepare a list of statements which cover in some detail the portion of the subject on which the pupils are to be examined. After such a list has been prepared, some of the statements can easily be changed so that they are false. The untruth of a statement should not be too obvious or it will be worthless for testing. Also statements should be selected which require an acquaintance with the subject in order to determine their truth or falsity.

2. In a true-false examination some of the statements should be true and some false, and the number of true statements should approximate the number of false statements. They should be arranged so that there is no regular sequence between true statements and false statements.

3. Since the pupil can give his responses very quickly, the examination should consist of not less than 50 statements. A true-false examination of 100 statements can be given in the time usually devoted to an ordinary examination.

4. The examination should be mimeographed or printed so that each pupil will have a copy. He may give his answers in the margins of the sheets, or, if it is desired to use the same set of papers with another group of pupils, he may be given a sheet of paper on which there are numbered blanks. The pupils will then be asked to record in the blanks their answers to the corresponding exercises. A less desirable plan, which may be followed when it is not possible to secure mimeographed copies of the examination, is to read the statements to the pupils and have them record their answers in numbered blanks. The disadvantage of this plan is that

---

[1] Two sample true-false examinations are given in the appendix.

the pupils do not have a satisfactory opportunity to study the statements. Also the class may give some indication of the answer if a statement appeals to them as being ridiculous.

5. The pupils should be given specific directions in regard to answering exercises about which they are uncertain. One writer[2] has suggested that the pupils be instructed to guess when uncertain concerning the truth or falsity of the statement. Another writer[3] who has used this type of examination instructed the pupils as follows: "First, go through the list quickly and mark all that you know for certain, then go back and study out the harder ones. Do not guess; the chances are against you on guessing. Don't endanger your score by gambling on those questions about which you know nothing." This second procedure is probably the better.

The scoring of a true-false examination. Since only two responses are possible, it is obvious that a pupil may give a correct response as the result of chance. In order to take this possibility into account, a pupil's score on an examination of this type is the number of exercises answered correctly minus the number answered incorrectly. Exercises not attempted are not counted.

b. "Yes" and "no" exercises. These exercises are just what their name implies. Each is answered correctly by "yes" or "no." No other answers are appropriate. The administration and scoring are similar to those of the "true-false" exercises, of which they may be considered a special type.

In constructing "yes" and "no" exercises care must be taken to avoid questions for which either of the two possible answers is absurd. In such case, obviously, a correct answer will not furnish satisfactory evidence of the pupil's achievement in a school subject. Approximately the same number of exercises should call for the answer of "yes" as for "no," and they should be arranged in no regular order of sequence.

c. **Recognition exercises.** Exercises in which the pupil is asked to choose from a number of proposed answers have also been used to make the scoring objective. This type of question has been called the "recognition exercise." It has been used in standardized silent reading tests and in a number of our group intelligence tests, and may be illustrated by the following:

"The first president of the United States was: Christopher Columbus, Benjamin Franklin, George Washington, Thomas Jefferson."

---

[2] McCall, W. A. "A new kind of school examination," Journal of Educational Research, 1: 33-46, January, 1920.

[3] Wood, Ben D. "The measurement of college work," Educational Administration and Supervision, 7: 301-34, September, 1921.

The correct answer may be indicated by an underline or by some other mark. If only one of the answers proposed may be considered correct the marking of such exercises will be highly objective.

Directions for constructing a recognition examination.[4] In constructing this type of examination none of the proposed answers should be too obviously incorrect. An exercise can yield an indication of a pupil's achievement only when he is forced to exercise judgment in determining which of the proposed answers is suitable. For example, the illustrative exercise given above would be practically worthless for testing purposes if all the names, except that of George Washington, were of persons living today or of persons who had had no connection with our national life. In applying this type of exercise to the field of arithmetic the proposed answers should include erroneous answers which pupils are inclined to give. For example, if the exercise called for the quotient of two fractions, one of the proposed answers should be the product of the fractions and another should be their sum, and perhaps another should be the fraction obtained by taking the sum of the numerator as a new numerator and the sum of the denominator for a new denominator. When the correct answer is included in a group of such answers as these, the pupil who does not know how to find the quotient of such fractions will be unable to determine the correct answer except as a matter of chance. On the other hand, if all of the answers except the correct one were integers or were so large that they were obviously incorrect, a bright pupil who knew nothing about division of fractions would be able to select the correct answer. The correct answer should not always be found in the same position; sometimes it should be first, sometimes last, and sometimes in an intermediate position. As in the case of the true-false examination, a recognition examination should consist of a large number of exercises.

Examinations of this type should be mimeographed or printed and a copy given to each pupil. He should be given definite instructions concerning the method of work to be followed. It is probably best to instruct him to work through the test rapidly, answering those exercises about which he is certain. He should then go back over the list and try the more difficult ones. Not fewer than four proposed answers should be given. When this is done the chances are slight that a pupil will give the correct answer by guessing. The pupils may be instructed to guess if they do not know, since the chance of success by guessing is slight. The pupil's score on an examination of this type may be taken as the number of exercises done correctly.

---

[4] A sample recognition examination is given in the appendix.

A somewhat unusual but interesting type of recognition exercise is that described as a "matching contest." In this a pupil is given two lists of statements, the first numbered 1, 2, 3, 4, 5, etc., the second marked A, B, C, D, E, etc. In the second list, there is a statement which corresponds in meaning to a statement in the first list and the pupil is to pair these statements, marking by the number of the first list the letter of the corresponding statement of the second. For example, in the exercises given below: by the date marked (5) 1898 we place the letter B to indicate the event for which that date is significant. It is difficult to construct such examinations so that they will require reasoning on the part of the student. Their most important use is in the elementary school for rapid drill in certain phases of some subjects, such as definitions in geography and grammar, etc. The following exercises, selected from the Spokane United States History Test, illustrate the use of such an examination in linking a certain date or person with the corresponding event.

1. 1846	A. Lincoln's Emancipation Proclamation.
2. 1865	B. Spanish American War.
3. 1863	C. Beginning of World War.
4. 1917	D. Declaration of Independence.
5. 1898	E. United States entered World War.
6. 1789.	F. Election of Washington as President.
7. 1792.	G. War with Mexico began.
8. 1776.	H. Invention of the cotton gin.
9. 1861.	I. Lee's surrender at Appomattox.
10. 1914.	J. Beginning of Civil War.

1. Foch	A. Destroyed Spanish fleet in Manila Bay.
2. Lincoln	B. Invented the telephone.
3. Fulton	C. Leading Confederate General.
4. Dewey	D. Wrote the Declaration of Independence.
5. Pershing	E. Invented the steamboat.
6. Bell	F. Commanded allied armies in the World War.
7. Edison	G. Was President during the Civil War.
8. Jefferson	H. Commanded American Forces in the World War.
9. Lee	I. Was Revolutionary patriot, author, and inventor.
10. Franklin	J. America's most famous inventor.

d. Completion exercises. Pupils may be asked to fill in the words omitted from statements. The following illustrates this type of exercise:

"Fill in the blanks. Revenue for paying the war debts of the states after the Revolutionary War was provided by a............and by ....due largely to............influence.

The slavery question in such states as should be carved from the Louisiana Territory was temporarily settled in..........by the......"

Directions for constructing completion exercises.[5] A completion exercise should be constructed so that no suggestion will be given of the correct words to be written in the blanks. Furthermore, the facts to be supplied should be important. The best plan is to prepare a list of important statements and principles covering the portion of the subject over which the pupils are to be examined and then from these statements to strike out a certain significant word or phrase. In every case, if it is possible, the words omitted should be such that only one answer will be correct. Since little writing is required of the pupils they may be asked to fill in as many as one hundred blanks.

The scoring of completion exercises is not as highly objective as in the two types mentioned above. Pupils will tend to write a variety of words in the blanks. Different words may have almost the same meaning, and frequently the scorer will be compelled to determine whether the meaning of some word is sufficiently near to that of the correct answer to justify giving the pupil credit for having answered the exercises correctly. However, by a careful selection of statements and of the omitted words, this subjectivity may be greatly minimized. For example, in the sentence, "The first Continental Congress was held in............," only one possible word can be correct. In using completion exercises it is necessary to provide each pupil with a mimeographed or printed copy of the examination. The pupil's score is the number of blanks filled in correctly.

e. Other advantages of the "new examination." Examinations in which true-false questions and the other types of exercises described on pages 29-34 are used have been called "new examinations." Such examinations have certain advantages in addition to increasing the objectivity of the marking of the papers. There will be a large saving of time both for pupil and teacher. The pupil is called upon to do little or no writing in giving his answers, and he is, therefore, able to respond to a large number of exercises. The scorer will have little or no occasion to exercise judgment as he will need only to note the brief responses given by the pupils. Thus, the labor of scoring will be greatly reduced. The saving of time in the giving and scoring will more than offset any additional time that may be expended in the construction of the "new examination." Another advantage is that the examination can be made more comprehensive. It is traditional for examinations to consist of ten questions. A few are limited to a smaller number, and only occasionally do we find examinations consisting of more than ten questions. The pupils cannot write upon a large number of questions in the time allowed. Thus, the scope of traditional examinations is necessarily narrow. We have stated that true-false examinations should include not fewer than 50 exercises. Examinations consisting of completion exercises or recognition exercises should have a corresponding length. Thus, the "new examination" may be made distinctly larger in scope. Children are apt, also, to be interested in the new examinations which are distinctly different from the usual type of school exercise, and which make an appeal somewhat in the nature of a game. In being relieved of much writing, which especially in the case of young children, amounts in itself to a laborious exercise, they are probably less fatigued, and are able to devote all their energies to the process of thinking.

---

[5] A sample completion examination is given in the Appendix.

f. Limitations of the "new examination." It does not appear likely that the "new examination," consisting of the type of exercises we have described, should replace entirely the traditional type of examination. The "new examination" cannot be used in mathematics, except to a limited extent. It cannot be used at all in English composition. The following questions taken from Hahn's Scale for Measuring the Ability of Children in History appear to require mental processes distinctly different from those for which the "new examination" calls.

"State points of similarity between the position of the United States in 1812 and her position in 1912.

"Arrange the following events in order of cause and effect: Force Bill, the Carpetbaggers, 15th Amendment, Negro Rule in some of the Southern States, Ku Klux Klan.

"Name the presidents of the United States since 1892."

Furthermore, it is likely that pupils would miss valuable experience and training if they were not asked at times to "compare," "explain," "discuss," "define," or "tell why." This is also true of questions in which they are asked to summarize material presented on a topic or to apply certain principles that have been presented. Hence, it is difficult to conceive of the "new examination" being a complete substitute for the traditional examination.[6]

B. Rules for marking examination papers. The marking of ordinary examination papers may be made more objective by following a few simple rules. It is advisable often to use the "sorting method." According to this procedure the papers as they are read are sorted into piles, the best ones being placed in the first pile, the next best in the second, etc. Five distributions will, in most cases, prove sufficient, but as the reading goes on, if papers are found which do not appear to belong in any of the piles, additional ones may be started. After all of the papers have been distributed they should be re-read, taking one pile at a time, and the papers in it compared with each other. If these papers do not possess approximately the same value changes in the sorting may be made. After the examiner has satisfied himself with reference to the sorting, "grades" may be assigned to the papers in the different piles. Instead of actually sorting the papers they may be marked with a symbol to indicate the distribution, i.e., mark the best papers 5, the next in quality 4, and so on. However, when this is done, the papers on which the same symbol has been placed should be brought together and re-read.

When a large number of papers are to be rated it is well to consider one question at a time. In doing this it is desirable to follow essentially the sorting method, or, at any rate, to assign marks to the answers as read. All of the papers should be read for the first question and a mark assigned it. Five seems a good number for approximately perfect answers, 3 for average, and 1 for very poor answers. The marks of 2 and 4 and even fractional numbers may be used as appropriate descriptions of quality. After the marks on this first question have been assigned, those papers which have been given the same marks may be brought together and re-read. The same procedure should be followed in reading each of the other questions. It is well to shuffle the papers in proceeding from one question to another so that the reader will not be influenced by previous marks. When the questions have all been read, the marks assigned to each paper may be totaled, and on this basis a final distribution of the papers made. Although this method may seem to require an almost prohibitive amount of time on the part of the scorer, the increased objectivity of the "grades" assigned may justify its adoption. Especially is it advised in those cases where emphasis needs to be placed on the accuracy of examination grades.

---

[6] In the appendix questions selected from actual examinations given in Illinois schools are listed. These questions in their demand for explanation, discussion, organization of material, exercise of judgment, etc., show that these traditional examinations call for a distinct type of ability not required by any form of the "new examination."

The marking of examination papers within a school system may be made more objective by formulating rules with reference to the credit to be given for certain types of answers. For example, uniform rules may be agreed upon with reference to the credit given for correct principle when the answer is wrong, the credit given for answers partially right, the penalty, if any, for misspellings, grammatical errors, and poor handwriting. In case the same examination is given by two or more teachers some specific rules may be formulated for the particular examination. If time permits, the objectivity of the marks may be increased by having the papers rated independently by two or more teachers and using the average of the marks assigned. A modification of this procedure is to have one teacher rate all of the answers to certain questions and another teacher the answers to other questions.

F. J. Kelly[7] describes an experiment which is indicative of the increase in the objectivity of the marking of examination papers when uniform rules are adopted. Six fifth-grade teachers gave the same examination in arithmetic to their pupils. Each teacher marked the papers for her own pupils but did not record the marks on the papers. The superintendent then asked a teacher, who was unusually systematic in marking examination papers, to prepare a set of rules to be followed in the marking of these papers. After she had done so, she marked all of the papers in accordance with this plan. Then the teachers who had first marked the papers marked them a second time following her plan. This provided two marks for each paper given by the classroom teacher, the first without following any systematic plan and the second given in accordance with the rules formulated. Each of these marks was compared with the mark given by the one teacher who marked all of the papers. In Table I, the six teachers are designated by the letters A, B, C, D, E, and F. The table is read as follows: When no rules were followed teacher A marked one paper 16 to 20 points lower than the "judge," one paper 7 points lower, two papers 4 points lower, two papers 2 points lower, agreed with the "judge" on one paper, etc. The differences between the marks given when the class-room teachers followed no rules and when they followed the rules formulated are very striking. In the first instance the marks assigned by the teacher agreed with those assigned by the "judge" in only 5.5 percent of the cases, while in the second instance they agreed in 63.5 percent of the

---

[7] Kelly, F. J. "Teachers' marks," Teachers College, Columbia University, Contributions to Education, No. 66, p. 83, 1914.

Table I. Distributions of Differences beween Two Sets of Teachers' Marks on Fifth-Grade Arithmetic Papers—First, without any Effort to Unify the Methods Used, and Second, by a Common Standard (after Kelly)

Range of Differences	Without Standard							With Standard						
	A	B	C	D	E	F	Total	A	B	C	D	E	F	Total
21 or more	....	....	....	....	2	....	2	....	....	....	....	....	....	....
16 to 20	1	....	....	....	1	1	3	....	....	....	....	....	....	....
15	....	....	....	....	....	2	2	....	....	....	....	....	....	....
14	....	....	....	....	....	1	1	....	....	....	....	....	....	....
13	....	....	....	....	1	2	3	....	....	....	....	....	....	....
12	....	1	....	....	1	....	2	....	....	....	....	....	....	....
11	....	....	1	....	1	2	4	....	....	1	....	....	....	1
10	....	....	....	....	....	1	1	....	1	....	....	....	....	1
9	....	1	....	....	2	1	4	....	....	....	....	....	....	....
8	....	....	....	1	3	1	5	....	....	....	....	....	....	....
7	1	1	....	1	1	1	5	....	....	....	1	....	....	1
6	....	2	....	....	1	1	4	....	....	....	....	....	....	....
5	....	1	2	1	1	2	7	....	....	....	....	....	....	....
4	2	2	2	1	1	2	10	1	....	....	....	....	1	2
3	....	4	2	1	2	2	11	....	....	1	1	1	....	3
2	2	2	1	1	1	1	8	4	1	1	3	7	1	17
1	....	5	4	3	2	4	18	2	3	4	5	1	1	16
0	1	4	4	1	1	1	12	22	30	16	16	29	26	139
1	2	5	2	2	2	1	14	5	....	2	2	1	3	13
2	6	1	3	2	3	1	16	1	1	3	....	....	....	5
3	9	....	2	....	2	....	13	....	2	2	1	....	1	6
4	5	1	4	1	5	1	17	....	2	3	3	....	....	8
5	2	3	2	2	1	....	10	....	1	1	2	....	....	4
6	1	1	....	3	2	....	7	....	....	1	1	....	....	2
7	....	1	1	6	1	....	9	....	....	....	....	....	....	....
8	....	2	1	2	....	1	6	....	....	....	....	....	....	....
9	1	....	1	2	....	....	4	....	....	....	....	....	....	....
10	1	....	....	1	....	1	3	....	....	....	....	....	....	....
11	....	1	1	....	....	....	2	....	....	....	....	....	....	....
12	1	....	....	1	....	1	3	....	....	....	....	....	....	....
13	....	1	....	....	1	1	3	....	....	....	....	....	....	....
14	....	....	....	....	....	1	1	....	....	....	....	....	....	....
15	....	....	1	1	....	....	2	....	....	....	1	....	....	1
16 to 20	....	2	....	....	....	....	2	....	....	....	....	....	....	....
21 or more	....	....	1	3	1	....	5	....	....	....	....	....	....	....
Totals	35	41	35	36	39	33	219	35	41	35	36	39	33	219
Medians	+3	0	+1	+6	−1	−4	+1							

cases. This indicates a very marked increase in the objectivity of the marking of the papers.

2. **Increasing the objectivity of the norms for translating examination scores into school marks.** The use of examination scores, when expressed in terms of percents, as school marks, introduces errors which are as serious as the errors arising from the subjectivity of the marking of examination papers. (See page 9). For standardized educational tests we have objective norms with which to compare the scores. Since a single examination is given to only relatively few pupils, it is not possible to secure objective norms by the procedure followed in standardizing educational tests. It is, however, possible to reduce greatly the errors in school marks arising from this source. The greatest assistance in this connection will be secured by recognizing explicitly that the scores assigned to examination papers and school marks are different. Occasionally they may coincide numerically, but they are fundamentally different. When this distinction is recognized it will be possible for a teacher to introduce objective elements into the norms that are used as a basis for translating examination scores into school marks.

A teacher should first determine whether or not his class is typical. The giving of a general intelligence test will be helpful in this connection. A distribution of their I. Q.'s may be considered a very reliable index of the composition of the group. If the median I. Q. of a class is below 100 the teacher may know that he has poor pupil material. If the median I. Q. is above 100 he may know that the class consists of better pupils than the average. If there is a relatively large number of low I. Q's it may be expected that there will be an unusually large number of low "grades." Thus, by means of the intelligence quotient and in other ways, the teacher may come to know the general status of his class.

The method to be followed in translating the examination scores into school marks may be illustrated as follows: Suppose that a set of examination papers has been rated in terms of percents. The score placed upon a paper simply describes the percent of the examination which, in the judgment of the scorer, the pupil has answered correctly. The scores should then be assembled in some such way as that indicated below:

			58		
			56	69	
		47	55	69	
	35	44	55	68	75
27	34	42	52	63	74
25	32	40	50	60	70

There are 23 pupils in this class. The scores have been arranged in six groups. It would be better to make five groups if the school uses a system of five marks. The median point score is 55. If the class is an average one this median score of 55 should be translated into the median or average "grade" which the school recognizes. If the "grades" are reported in terms of percents and the passing mark is 75 the average grade will usually be approximately 85. If the class is known to possess superior ability the median score of 55 should be translated into a higher "grade." On the other hand, if the class is known to be decidedly below average in ability, 55 should be translated into a lower grade, perhaps as low as 77 or 78. In an extreme case it might be translated even into the passing grade of 75. The translation of the median score into a grade furnishes a basis for translating the other scores. In general, the scores at the lower end of the distribution will be translated into "grades" below passing. The percent of pupils who receive such "grades" will vary widely with different classes.

There is a somewhat prevalent opinion that the normal probability curve fixes the percent of pupils who should receive "grades" below the passing mark. This is a mistaken notion. The normal probability curve tells us nothing concerning the percent of pupils who should receive any "grade." It is true that from a statistical point of view there are certain divisions of the curve which are convenient. If the base line of the curve is limited to a length equal to five times the standard deviation (5σ) and this distance is divided into intervals of 1σ and perpendiculars are erected at the division points, the percent of cases falling in each division of the curve will be 7, 24, 38, 24, and 7. It has been suggested that these numbers define the percent of pupils who should receive grades of A, B, C, D, and F, respectively. Similar proposals have been made for other plans of dividing the normal probability curve.

No plan of dividing the normal probability curve can claim to be distinctly superior to any other plan. One may accept, without being inconsistent, the assumption that the "grades" for a large group of pupils should conform to the normal probability curve, and at the same time refuse to accept any particular proposed specifications as to the percent of pupils who should receive each "grade." For example, a distribution of 50 percent A's, 25 percent B's, 10 percent C's, 9 percent D's, and 6 percent F's is not necessarily inconsistent with the assumption that accurate measures of achievement tend to form a normal distribution. However, in this case, it would be necessary to define a "grade" of A as being a "grade" which means that the pupil is above average ability or, in other words, that all pupils who are average or above are given a "grade" of A without any attempt to distinguish between their achievement. Thus a "grade" of A would represent a wide range of achievement. On the other hand, "grades" of B, C, D, and F represent narrow ranges of achievement. Thus, the different "grades" would not represent equal ranges of achievement. This, however, for certain purposes, may not be entirely undesirable.

The range of achievement which a "grade" shall represent, or, in other words, the percents of pupils who in the long run shall receive the different "grades," is a matter of school policy. This should be determined by the school. Perhaps the best way to define the range of achievement which a "grade" is to represent is in terms of the percent of pupils who, in the long run, will receive the "grade." Undoubtedly, at the present time, there is considerable variation from system to system in respect to the range of achievement that is represented by the different "grades." If a system can reach a definite agreement concerning the percent of pupils who should receive each mark a step will be taken in the direction of making objective the norms which are used in translating examination scores into school marks. If only five marks are used a school would probably not be far from the general practise if the percents of pupils receiving these "grades" were defined as 7, 24, 38, 24, and 7. However, in establishing this definition it should be distinctly recognized that it is done as a matter of school policy instead of being forced upon the school by the nature of the normal probability curve.

3. Securing closer agreement of examinations with educational objectives. One of the weaknesses of examinations set by teachers and other school officials is that questions are frequently asked which call for non-important facts or which are not in agreement with the emphasis that has been placed upon the course during the term. By creating a proper attitude toward examinations on the part of teachers much can be accomplished toward bringing the examinations into agreement with our educational objectives. One who sets an examination should try to have it include the more important points that have been taken up during the term, and should make a distinct effort to omit all points that are unimportant. In the case of a few school subjects we now have available a partial list of minimum essentials. One setting an examination may use such lists in determining the relative importance of different topics. Textbooks, also, in that they represent the judgment of experts concerning what topics are most important, are valuable guides in this connection. Those topics which are treated briefly or are obviously assigned to a place of minor importance should not be included in an examination.

CHAPTER IV

SUMMARY

1. Probably the most prominent function of the written examination is that it provides an opportunity for learning, both in the preparation for the examination and in the actual writing of it. The pupil who is not required to take examinations is missing an important part of his educational opportunities. The importance of this function, however, varies with the type of examination. Examinations which call merely for facts or in which the pupil is not required to formulate an answer consisting of several sentences offer only a limited educational opportunity. Incidentally it may be noted that this is one weakness of the "new examination."

2. A second function of written examinations is the measurement of the achievements of pupils. As we have pointed out, measurements are essential to the organization and administration of schools as well as to class-room instruction, and examinations do secure a type of information relative to the achievements of pupils which can be secured in no other way. Although written examinations possess these functions it should be remembered that it is only when examinations are properly used and not abused that the fulfillment of these functions may be expected.

3. The written examination furnishes an effective motive. It is true that some criticisms may be offered against the use of the examination to motivate the work of the school, but this use, unless carried to an extreme, is probably not harmful.

4. Final examinations have a positive value in that they call for intensive review and organization of the content of the course on the part of the student. This emphasizes an important phase of learning.

5. In certain school subjects standardized objective tests are superior to written examinations set by the teacher as instruments for measuring the achievements of pupils. Under certain conditions teachers' estimates and daily grades may be more accurate measures of achievements than can be secured by means of written examinations. However, in view of the other functions of written examinations, we are justified in asserting that in general there is no satisfactory substitute for them. The written examination is more important in some school subjects than in others, and frequently it furnishes the teacher with a type of measurement of the achievements of pupils that cannot be secured by any other

means. In addition, it is necessary to recognize that certain types of written examinations fulfill other important functions.

6. In arriving at final "grades," teachers' estimates, daily "grades," and "grades" earned on short quizzes should be combined with examination "grades." No general rule can be stated concerning the weight that should be given to each of these types of measure in arriving at final "grades." Under certain conditions and in some school subjects the final examination should receive considerable weight, perhaps as much as 50 or 60 percent. In other cases it should be given a much smaller weight.

7. Teachers should give more attention to the formulation of examination questions. Relatively unimportant topics or those which have not received emphasis during the term should be omitted. Ambiguous statements should be avoided. The questions also should be varied so that they demand different types of mental ability, memory, reasoning, organization, etc. Pupils should be given directions concerning methods of work.

8. Examinations should be long enough so that practically no pupils will finish. In this way a measure of rate of work will also be obtained.

9. The examination should be considered a part of the regular school work. A written examination has a unique educational value. Pupils who are excused from it or who are not kept occupied throughout the entire period are being deprived of an important educational opportunity.

10. A systematic procedure should be followed in marking the examination papers. Explicit rules should be formulated relative to the effect of misspelling, poor English, poor handwriting, and so forth, upon the "grade" assigned to a paper. The rule should cover also credit for correct principle and partial credit for answers partially right. The "sorting method" of marking is recommended.

11. The difference between scores and "grades" should be recognized. Objective factors should be introduced into the norms used for translating the scores into "grades."

12. The "new examination" has many advantages, among which are its increased objectivity of marking and the economy of time for both pupils and teachers. The "new examination" will undoubtedly prove useful, but its limitations must not be overlooked.

13. Traditional examinations call for the functioning of distinct types of mental ability not demanded to so great a degree in any other kind of school work, and should not, especially in the case of final examinations, be abolished, nor replaced entirely by any other form of school exercise. Teachers and pupils, more and more, should be impressed with their unique educational value.

APPENDIX

(The following "new examinations" are given for purposes of illustration. They may include several exercises which will prove unsuitable when given to pupils.)

TRUE-FALSE EXAMINATION IN PHYSIOLOGY

Prepared by
Bureau of Educational Research
University of Illinois

Name..Boy or Girl...

Age last birthday........................Next birthday will be...19........

Grade........................ Date................................ City................................ State.................................

School...Teacher...

Below you will find a number of statements. Some of these statements are true, others are not true. Read each statement carefully, then if it is true mark a plus (+) in the column to the right of the sentence. If the statement is not true mark a minus (−) in the column to the right.

Answers to be written here

EXAMPLES

Read the statement below very carefully.

1. **Fats will form a lasting mixture with water.**

This is not a true statement so you will place a minus (−) sign in the column
Now read the second sentence.

2. **The layer of fat just beneath the skin is more than one-tenth of an inch thick.**

This is a true statement so you will mark a plus (+) sign in the column........
Now read the third sentence.

3. **The union of oxygen with any substance produces heat.**

This is a true statement so you will mark a plus (+) sign in the column........
Now read the fourth sentence.

4. **Nitrogen constitutes only one-fifth of the volume of the air.**

This is a false statement so you will mark a minus (−) sign in the column....

PHYSIOLOGY

Answers to be written here

1. The kidneys vary 12 inches to 16 inches in length.
2. The external poisoning of the skin by poison ivy or sumac never results seriously.
3. A person having a good mind must necessarily have a large brain.
4. Color blindness is more prominent in men than women.
5. Plenty of fluids should be drunk at the time of eating solid food.
6. Bones are composed of animal and mineral matter.
7. The nails are hardened outer skin or epidermis.
8. The use of alcohol increases the tendency to commit crime.
9. A full grown person contains about six quarts of blood.
10. The brain is almost perfectly spherical in shape.
11. The kidneys are almost perfectly round.
12. All animals are made up of cells.
13. Substances, like glass, which permit rays of light to pass through them readily are said to be opaque.
14. The sense organs of smell are located in the lining of the cavity of the nose.
15. The skin is composed of two layers of tissue.
16. The end organs for taste occur in the mucuous membrane of the tongue.
17. To extinguish the burning clothing of a person, it is necessary to wrap him in something to exclude the air.
18. All of the interior of the spinal cord is filled with gray matter containing nerve cells.
19. The great difference in the complexion of persons is due largely to the pigment lying in the epidermis.
20. Cancer is caused by germs growing in the tissue.
21. Diptheria can be controlled by the use of Diptheria antitoxin.
22. The brain is separated into two parts or hemispheres by a great longitudinal fissure.
23. When oxygen is separated from other substances the process is called oxidation.

Answers to be written here

24. Infectious diseases are due to changed methods of work and growth on the part of cells in certain regions of the body.

25. The use of alcoholic beverages builds up the body and makes the muscles stronger.

26. The great majority of grown people have been infected with tuberculosis germs.

27. The sense organs are the terminations of the sensory nerves serving to carry impressions to the spinal cord or brain.

28. Farsightedness is often caused by a blow on the eye.

29. An antiseptic is a substance which merely restrains the germs from growing.

30. The brain is in communication with the rest of the body by means of nerves.

31. The cerebrum is the path of communication between the nerves supplying the arms, trunk, legs, and brain.

32. The chief function of muscles is to hold up the body.

33. All milk contains bacteria.

34. The alcohol used in drinks is produced by the growth of yeast in a liquid containing sugar.

35. Our blood contains white corpuscles which destroy disease germs.

36. More people die daily from diptheria than from tuberculosis.

37. The use of tobacco increases the strength of the muscles.

38. The use of tobacco makes the nerve cells function more keenly.

39. The chewing of dry bread aids the digestion as much as the use of gum.

40. Air is composed chiefly of two gases, oxygen and nitrogen.

41. The first step in treating a person who has been poisoned is to give an emetic.

42. Light is produced by waves of a substance called ether.

43. Non-infectious diseases are caused by small plants or animals called parasites feeding upon the human body.

44. Alcoholic beverages have great value in curing disease.

45. A drink of alcoholic beverage in the winter time causes a man's body to become warm.

	Answers to be written here
46. Each portion of the brain has its own definite work to perform.	
47. Fainting is caused by an over-sufficient supply of blood being sent to the brain.	
48. The spinal cord may act independently of the brain and produce many of the muscular movements necessary in routine work.	
49. The germs of typhoid fever usually gain access to the body by being breathed in with air.	
50. Narcotics are substances which cause any organs of the body to act more vigorously than is their custom.	

Directions to teachers: After the four examples have been studied by the pupils, read the following directions to them: "On the next page you will find a number of statements similar to the ones you have just read. You are to place a plus sign or a minus sign in the column to the right of each statement just as has been done on the first page. Mark all of the statements that you are sure you can answer correctly. If you find a statement that you are not sure you can answer correctly, study it carefully and then mark the answer you think will be correct. If you find a statement you know nothing about, make no attempt to mark it, as guessing counts heavily against you. You will have 25 minutes for the test. I shall expect you to stop promptly and turn your folders face down on the desk when I tell you to do so. Ready-Go."

In computing the score of each pupil on a test subtract the total number of wrong answers from the total number of right answers. Such scores are called "point-scores." In interpreting them it is advisable to form a distribution which will show how many pupils received each score. From the distribution it is possible to work out a basis for translating the point scores into the usual kind of school marks.

Note: The "Directions to teachers" given above would not appear on the usual printed examinations. They are placed here for the convenience of teachers.

TRUE-FALSE EXAMINATION IN HISTORY AND CIVIL GOVERNMENT

Prepared by
Bureau of Educational Research
University of Illinois

Name..Boy or Girl..

Age last birthday........................Next birthday will be..19........

Grade........................Date...............................City...................................State..................................

School...Teacher..

Below you will find a number of statements. Some of these statements are true, others are not true. Read each statement carefully, then if it is true mark a plus (+) in the column to the right of the sentence. If the statement is not true mark a minus (—) in the column to the right.

Directions to Teachers: Unless the students are familiar with the True-false Examination, explain the procedure to them by using four examples as given in the Examination in Physiology, Appendix, p. 44; Time allowance for this examination, 37 minutes. See also **Directions to teachers**, Appendix, p. 47.

Answers to be written here

1. During Jackson's time the national convention method of nominating presidential candidates was introduced.
2. The Constitution was gladly accepted by all of the people.
3. Jefferson was the second president of the U. S.
4. John C. Calhoun was a strong advocate of the doctrine of "Nullification."
5. Throughout the Revolutionary War the American navy was superior to that of England.
6. The reconstruction period following the World War was a very troublesome one in the U. S.
7. During the first two years of the Civil War the confederate troops won the majority of the victories.
8. England removed all the taxes against which the colonists rebelled.

Answers to be written here

9. Scientific farming in the U. S. was little thought of before 1860.

10. France entered the World War to save Russia.

11. Robert E. Lee, leader of the Confederate forces, was one of the greatest soldiers of the Civil War.

12. In the war of 1793 between France and England, the U. S. sided with France.

13. In the battle with the English warship "Serapis," the flagship of Paul Jones, the "Bon Homme Richard," was completely destroyed.

14. President Hayes was elected over his rival candidate by an overwhelming majority.

15. The Federal Reserve Act of 1913 exercises an important effect in preventing money crises.

16. During Monroe's presidency there was more internal strife and turmoil than during any previous period since the Revolution.

17. John Brown's Raid at Harper's Ferry brought many non-slave-holders to the side of the slave-holders.

18. Russia entered the World War to protect Serbia.

19. The Lincoln-Douglas debates won Lincoln the Illinois senatorship over Douglas.

20. In ten years the tolls and revenues from the Erie Canal repaid the entire amount spent during the eight years of its construction.

21. During the period of reconstruction following the Civil War, Congress passed laws prohibiting the negroes from voting.

22. The issue of Wilson's second campaign was the keeping of the U. S. out of war.

23. At the end of the Spanish-American war, Spain turned Cuba over to the U. S. as a colony and paid us $20,000,000.

24. Many important inventions were made between 1830-1860.

25. Washington Irving was the first great American poet.

26. During the revolution the population of the United States increased.

27. John J. Pershing, U. S. General, was put in supreme command of all of the allied forces during the World War.

28. Andrew Jackson was one of the most cultured of our presidents.

29. The Missouri Compromise prevented slavery in all of the Louisiana territory north of 36° 30', except in Missouri.

Answers to be written here

30. At the battle of Manila Bay, a German fleet would have helped the Spaniards had not the Admiral of a British fleet warned the Germans not to interfere.

31. The Dred Scott Decision meant that negroes could not become citizens.

32. England entered the World War because the German fleet bombarded English coastal towns.

33. In tilling the soil, the farmers at first were afraid to use a cast-iron plow, saying that it poisoned the soil.

34. The English government sent American cotton manufacturers copies of the spinning jenny soon after its invention.

35. In the final battle of the Texas rebellion against Mexico, the Mexicans were outnumbered two to one.

36. The early colonists believed in witchcraft.

37. General Braddock's army saved Washington and his Virginian troops from an Indian massacre.

38. Georgia was populated by people who had been thrown into jail for debt.

39. William Pitt, as prime minister of England, was a great aid to the colonists in driving the French from the Atlantic coastal region.

40. In the war of 1812, the Americans won many brilliant victories on the sea.

41. Lincoln was killed in a battle during the Civil War.

42. Washington belonged to the Federalist party.

43. The Referendum is a process of obtaining legislation which the people desire, in spite of an unwilling legislature.

44. From 1830-1860 the population of the U. S. doubled.

45. John Adams was the first president to be inaugurated in the city of Washington.

46. During the colonial period books were plentiful.

47. The British encouraged the Indians to kill American settlers who moved into the Northwest Territory.

48. Manufacturing was the important industry from the time of the Revolution down to Jackson's time.

49. The battle of Gettysburg was the only battle of the Civil War fought on northern soil.

Answers to be written here

50. The Initiative is a process of preventing legislation which a legislature would impose upon the people.

51. James Fenimore Cooper was the first American author to receive recognition in Europe.

52. The South, after all of the states had seceded, had more men for soldiers than the North.

53. England was not as kind to her colonies as the other European nations.

54. The battle ship Maine was sunk in the harbor of Havana by a torpedo fired from a Spanish warship.

55. At the close of the Mexican war, Mexico was forced to pay the U. S. $100,000,000 and to pay claims of $3,500,000 which American citizens had against her.

56. The presidential message sometimes has great influence upon Congress.

57. During the administration of President Hayes much bitter feeling arose between the North and the South.

58. Lincoln accepted the presidency with the determination to forever wipe out slavery.

59. The annexation of Texas led directly to war with Mexico.

60. William Lloyd Garrison was a violent anti-abolitionist.

61. During the period from 1750-1800 the steamboat was the main means of travel where streams were available.

62. Many northern people opposed Lincoln's reelection.

63. John Adams was the second president of the U. S.

64. One result of the war of 1812 was to stimulate manufacturing in New England.

65. Madison was very much opposed to Jefferson's policies.

66. The battle of Gettysburg was the turning point in the Civil War.

67. The spoils system has been very beneficial to this country.

68. Jay's treaty with England was much opposed in the U. S.

69. The underground railroad was a railroad used to carry cotton from the south to the mills of New England.

70. John Quincy Adams possessed the characteristics of an ideal president.

71. The beginning of the westward movement was during the administration of John Quincy Adams.

72. Washington was inaugurated in New York City.

73. During the latter part of the 18th century the U. S. was carrying on trade with China.

74. The war with Mexico lasted over three years.

75. Lincoln declared war on the southern states to make them give up slavery.

TRUE-FALSE EXAMINATION IN GEOGRAPHY

Prepared by
Bureau of Educational Research
University of Illinois

Name..Boy or girl..

Age last birthday......................Next birthday will be..19.......

Grade......................Date..............................City..............................State.................................

School...Teacher.... ....

Below you will find a number of statements. Some of these statements are true, others are not true. Read each statement carefully, then if it is true mark a plus (+) in the column to the right of the sentence. If the statement is not true mark a minus (−) in the column to the right.

Directions to teachers: Unless the students are familiar with the True-False Examination, explain the procedure to them by using four examples as given in the Examination in Physiology, Appendix, p. 44. Time allowance for this examination, 40 minutes. See also **Directions to teachers,** Appendix, p. 47

	Answers to be written here
1. Agriculture is the principal industry of France.	
2. Every portion of South America has a very warm climate during all seasons of the year.	
3. The lumber resources of India have been little developed.	
4. Very little rain falls in the vicinity of the equator.	
5. In the forests of South America are found many fierce, wild animals.	
6. China is governed by a republican form of government.	
7. Trees prevent rain water from running rapidly off the land.	
8. The main industry of Australia is manufacturing.	
9. Russia in Europe has many good harbors.	
10. Many Europeans have settled in Central Africa.	
11. Switzerland has large coal and mineral deposits.	
12. Argentina is one of the more backward of South American countries.	

Answers to be written here

13. There is not a great variation in the amount of rainfall which is received by different sections of Europe in a year.

14. People who live in a warm climate are not as industrious as those who live in a temperate climate.

15. At the time of the arrival of Columbus the natives of South America were all in the lower stages of barbarism.

16. Manufacturing is an important industry in Germany.

17. When Columbus discovered South America it was inhabited by both red and white men.

18. London is the largest city in the world.

19. The coast of Australia is very regular.

20. The population of Great Britain is nearly one-half as great as that of the entire U. S.

21. Africa has two deserts, one north and one south of the equator.

22. Belgium, though smaller, has more people than Holland.

23. France produces more coal than the British Isles.

24. Much of the Netherlands lies below sea level.

25. The work done by the Great Ice Sheet was a hindrance to the further development of our country.

26. Very little mining is carried on in Spain and Portugal.

27. China has the largest number of inhabitants of any country in the world.

28. Before the time of Christ, Greece was one of the most backward and barbaric countries of the world.

29. Asia has much the same climate throughout.

30. Corn is the principal food of the Japanese people.

31. The Brazilian forests produce many valuable products.

32. The Mississippi Valley is the most extensive farming section in the United States.

33. Italy has many very valuable mineral deposits.

34. The plains of Russia are populated by many different peoples.

35. Agriculture is the principal industry of Belgium.

36. In early times the Chinese led the world in the matter of civilization.

Answers to be written here

37. In early historical times, the Mississippi river was longer than it is now.

38. The people of Spain and Portugal are very backward.

39. The majority of the Asiatic people belong to the yellow race.

40. Europe has the most regular coast of all the continents.

41. All of Australia has a very heavy rainfall.

42. Chile is one of the great copper-producing countries of the world.

43. Manufacturing is the principal industry of South Africa.

44. China has more inhabitants than all European countries taken together.

45. Vegetable and fruit growing is extensively followed on the Coastal Plains.

46. Brazil is smaller than the United States.

47. Agriculture is the principal industry of Switzerland.

48. All of the territory covered by Spain and Portugal has abundant rainfall.

49. The French, attracted by the discovery of rich deposits of gold and silver, seized almost all of South America.

50. Our present population is over 150,000,000.

51. The principal industry of Venezuela is mining.

52. Very few forests are found in European Russia.

53. Some parts of the German Empire are too arid for agriculture.

54. South America is quite like North America in its surface features.

55. Many manufacturing industries have been established along the western coast of the United States.

56. Africa is governed by many different nations.

57. Chile is one of the least progressive nations in S. A.

58. Great Britain has more manufacturing than any other nation.

59. Europe should not be called a continent.

60. Very little agriculture is carried on in Argentina.

61. Much of the coast of South America has been sinking.

62. Australia is larger than the U. S.

63. Irrigation is little practised in the western states.

64. Great Britain has a greater foreign trade than any other nation of the earth.

Answers to be written here

65. Very little manufacturing is carried on in Greece.

66. Little agriculture is carried on in Uruguay and Paraguay.

67. The discovery of gold in Australia led to the development of that country's resources.

68. Commerce in Holland is highly developed.

69. Irrigation is extensively developed in India.

70. Agriculture is the principal industry in Italy.

71. Most of Russia is a level plain.

72. Egypt is now governed by the Turks.

73. The Chinese cities are very crowded.

74. Most of the Asiatic people are either Brahmans or Buddhists.

75. The government of Great Britain is a limited monarchy.

76. The Nile river flows across the desert.

77. The Delaware Water Gap is in Pennsylvania.

78. Commercially, Italy occupies a very favorable position.

79. Many more animals have their homes on land than in the sea.

80. The principal occupation of Greece is coal mining.

81. Grain farming is the principal industry of Rumania and Bulgaria.

82. More than half of Europe is an extensive lowland.

83. More than one-third of our population is made up of foreigners and and negroes.

84. Animal life is found in great abundance throughout South America.

85. Ireland is mainly a country of farms.

86. The roads and railways of Switzerland are very poor.

87. The population of North America exceeds 200,000,000 people.

88. The Sahara desert is practically as large as the U. S.

89. Wheat is the main crop in the Valley of the Red River of the North.

90. Mount Vesuvius, in Italy, is constantly in eruption.

91. In some places the Amazon river is several miles wide.

92. The Japanese people have been very quick to learn the lessons of western civilization.

Answers to be written here

93. Uruguay and Paraguay are very mountainous countries.

94. Very little manufacturing is carried on in the Middle Atlantic States.

95. The Tropical Andean Countries are among the most important mineral producing regions of the world.

96. The principal food of the Chinese is corn.

97. All of Africa is populated by the negroes.

98. Much of the surface of Japan is mountainous.

99. The people of Asia have always been very backward.

100. The government of Japan is an absolute monarchy.

COMPLETION EXAMINATION IN AMERICAN GOVERNMENT

Prepared by
Bureau of Educational Research
University of Illinois

Name...Boy or Girl...

Age last birthday.......................Next birthday will be...19........

Grade.......................Date...............................City...............................State..

School...Teacher...

Below you will find a number of statements. In each statement one or more important words have been omitted. Each blank in the sentence shows where a word has been left out. Read each statement carefully, then write in the blank the word which completes the meaning of the statement. You will be allowed 15 minutes for the test.

1. The primary purpose for which government exists is the...............................of our lives and property.
2. Citizenship may be acquired by...............................in this country or by a process of ....for natives of other lands.
3. Our national government derives its authority from the...............................of the United States through our national.....................................
4. The legislative power granted to the national government is vested in a Congress of ....houses, the smaller of which is called the...............................and the larger the...............................
5. The execution of the laws made by...............................is intrusted to the........................... of the United States.
6. All judges connected with the national courts are appointed for life with the consent of...
7. Most of the candidates for office which are filled by popular vote are nominated directly in..
8. The Fifteenth Amendment of the United States Constitution prevents the states from denying to citizens the right to vote on account of.., ..., or previous condition of...

9. Practically all of our law-making bodies are made up of..chosen for short terms from....................................into which the states, counties and cities are divided.

10. The first permanent English settlements in America were made in what is now the state of...

11. In a county, the records of the county board and other official papers are preserved by the county..

12. All cities are public corporations created under...............................municipal laws.

13. Every incorporated city obtains from the.......................................government a............... under which it may elect its officials and conduct its business.

14. Civil service employees may be removed from service only for...

15. The power of impeaching a state officer is given to the...

16. The..is by far the most prominent and powerful executive official in the state. Very................................state officers are appointed by him or are responsible to him.

17. All important officials connected with the executive or judicial service of the United States may be removed by.......................................through the lower house of Congress and by.......................................in the senate.

18. Far more property is destroyed by..than by all other agencies.

19. There is no task of state and local government which outranks in importance that of providing an..education at public expense.

20. All rivers and canals within a single state are controlled by the.. in which they are located.

21. Most of the revenue for state and local governments is secured by a.............................. on............................

22. A state................................is the fundamental law which the people of the state have arranged for their government and protection.

23. A state constitution can be changed by means of an..

24. The three-fifths compromise provided that five...............................should be counted as equal to three.......................................when reckoning the...............................for either direct taxation or representation.

RECOGNITION EXAMINATION IN ALGEBRA

Prepared by
Bureau of Educational Research
University of Illinois

Name..Boy or girl..

Age last birthday.......................Next birthday will be..19....

Grade.......................Date...............................City...............................State..

School...Teacher..

Below you will find a number of statements. In each statement a word or number has been omitted. At the close of the statement several words or numbers have been given. One of these is the correct answer. Select the word or number which you think is correct and draw a line under it. Most, if not all, of the examples can be solved by mental calculation. If any figuring is necessary, work on the margin of the page. You will be allowed 17 minutes for the test.

1. Numbers that are represented by letters are called..................................numbers.
   substituted—literal

2. When two or more letters are multiplied together each is called a......................................
   of the product. factor—coefficient

3. If a man rides a certain distance in 10 hours, in h hours he rides.......................................
   $10h; \frac{h}{10}; \frac{10}{h}$

4. The statement $2x+5=29$ is called an.......................... identity—equation.

5. If 16 is substracted from three times a certain number the result is 110. The number
   is...................... 36 2/3; 31 1/3; 42.

6. A number which is a factor of two or more numbers is called a................................factor.
   common—equal

7. If there are two equal factors of a number, either is called the............................of the
   number. square root—common factor.

8. To multiply algebraic fractions take the.......................................of the numerators for a
   new numerator and the product of the denominators for a new denominator.
   sum—product

9. A fraction whose numerator or denominator (or both) contains fractions is called a ....fraction. multiple—complex

10. A ....is a statement of a fact which is to be proved. theorem—axiom.

11. The name given the + sign is.................................. negative—positive

12. To find the sum of two numbers whose signs are opposite, take their.............................. regarding each as positive, and prefix the sign of the larger number to the answer. sum—difference—product

13. Whenever a number occurs without a sign, the..............................sign is to be understood. X; +; −

14. The number denoting the power of a term is called the.. prefix—exponent

15. If $a=2$, $b=-3$ and $c=-5$ then $\frac{2abc}{a}=$.......................... -6; -2; 30

16. In adding like terms add the coefficients for the new coefficient and.............................. it by the common factor. multiply—divide.

17. An expression which contains more than one term is called a... monomial—polynomial

18. If the length of a rectangle = 4 feet more than twice the width, the perimeter = 56 feet. The length = ...feet. 8—12—16

19. Any term may be transposed from one side of an equation to the other, provided its ....is changed. sign—value.

20. Any equation which contains no higher power of the unknown letter than the first is called a.....................................equation. radical—simple.

21. The exponent of the product of two powers of the same number is equal to the ....of the exponents of the factors. product—sum

22. To raise the product of two numbers to any power, raise the numbers separately to that power and take their...................................... product—sum

23. The square of any two numbers is equal to the square of the first number......................... twice the product of the two plus the square of the second number. plus—minus

24. A 20 foot ladder rests against a building, the bottom of the ladder being 12 feet from the cellar wall. The top is........feet from the ground. 8-16

25. In division, the sign of the quotient is..............................whenever the dividend and divisor have like signs. −; +

26. In finding the quotient of two powers of the same number the exponent of the quotient is equal to the exponent of the dividend...............................by that of the divisor. increased—diminished.

27. $(3x^2-2x-1)\div(x-1)=$........................ $3x+1$; $3x-1$.

28. A factor which has no factor except itself and unity is called a...............................factor. prime—multiple

29. The product of all the common prime factors of two or more numbers or expressions is called their......................................common factor. highest—lowest.

30. If one number is exactly divisible by another, the first is called a...............................of the second. Divisor—multiple

31. In algebraic fractions the dividend is called the...
denominator—numerator

EXAMPLES OF TRADITIONAL EXAMINATIONS

At the request of the Director of the Bureau of Educational Research, a number of superintendents and high-school principals in Illinois sent copies of final examinations which were being given in their schools during the year 1921-1922. The following list of questions has been selected from the large number submitted. These questions, although the selection was made on the basis of the judgment of one person only, seem among the best examples of the traditional type of examination. In that they demand explanation, discussion, organization of material, exercise of judgment, etc., on the part of the pupil, they call for a distinct type of ability not required by any form of the "new examination." In making this selection the different school subjects and the different grades both of the elementary and of the high school have been represented to as great a degree as possible. In the lower grades of the elementary school the majority of the questions submitted were based upon memory and could easily have been changed into some form of the "new examination." For that reason, no examinations from those grades are included in this selection.

HISTORY

(Seventh grade)

1. What was the purpose of Columbus' voyage and its result?
2. What did ten of the following explore or discover: Cabot, Balboa, Ponce de Leon, Magellan, Coronado, De Soto, Drake, Hudson, Cartier, Champlain, LaSalle.
3. What people settled Jamestown, Virginia? Discuss one of the following topics in connection with Virginia history.
   The Starving Time, Individual Ownership, Tobacco Raising.
4. Tell by what class of people and why was each of the following states settled: Maryland, Carolina, Georgia.
5. Tell the story of William Penn, naming the colony he founded.
6. Why did the Pilgrims come to America? Where did they land? Write of their first winter in America and their relations with the Indians.
7. Tell what happened in 1492, 1607, 1619, 1620, from 1519 to 1522.
8. For what nation and on what errand did Joliet come to the Illinois country? Why did Marquette come with him?
9. Tell the story of Starved Rock. What tribes of Indians were connected with it and how?
10. Write an item of historical interest about each of five: Tonti, Stuyvesant, Oglethorpe, Baltimore, Bacon, Bradford.
11. Write a short paragraph about two of the following:
    1. Illinois pioneers.
    2. Illinois rangers.
    3. Block houses.
    4. Keel boats.

HISTORY

(Seventh grade)

1. What were the results of the Revolutionary War?
2. Explain in what ways Congress was weak under the Articles of Confederation.
3. Who was Lafayette? What did he do for American liberty? Why?
4. What was the Ordinance of 1787?
5. A convention of delegates from the states was called to meet in Philadelphia in May, 1787. Why?
6. Who was the first President of the U. S. under the Constitution? When and where was he inaugurated?
7. Name and tell how important inventions have helped the progress of the U. S.
8. Why was the purchase of Louisiana an important event for the U. S?
9. What was the result of the War of 1812?
10. Tell something of the work of the Humanitarians and the establishment of the free elementary schools.

HISTORY

(Eighth grade)

1. Discuss five powers or duties of Congress.
2. Name the officers of the President's cabinet and a duty of each.
3. Explain the need of a survey system. Make a diagram showing baseline, principal meridian, township lines, range of townships. Locate Twp. 2 N. R. 3 E of P. M.
4. How were ten of the following connected with the Civil War: Stonewall Jackson, Major Anderson, Robert E. Lee, Jefferson Davis, McClellan, Hooker, Grant, Sherman, Emancipation Proclamation, Hammering Campaign, Gettysburg, and Appomattox Court House.
5. Tell the location, time, inventor, and importance of the Atlantic Cable.
6. Tell of two good results of the Civil Service Reform.
7. Name and discuss two famous laws we have studied this semester.
8. Write a brief paragraph discussing the importances of the Pan-American Congresses.
9. Give the time, place, purpose and importance of an exposition studied this semester.
10. State two causes and two results of the Spanish-American War.
11. Write a statement about each in connection with the World War: Autocracy; "der Tag;' submarine; Lusitania; armistice.

HISTORY

(High school)

1. Discuss the work of Spain in exploration—naming five important explorers.
2. State the Mercantile Theory of trade and explain its effects during the American colonial period.

3. State five defects in the Articles of Confederation. How were these defects remedied in the Constitution?
4. What was the Northwest Ordinance of 1787? Why important?
5. Explain Alexander Hamilton's policy on the U. S. debt.
6. Give the history of the Nullification Controversy of 1828-33.
7. Give the history of the election of 1823.
8. Identify the following: Gallatin; Oglethorpe; DeWitt Clinton; Thos. Paine; Stephen Decatur.
9. Explain three important results of the war of 1812.
10. Discuss the political platform and policies of President Jefferson.

ANCIENT HISTORY

(High school)

1. (a) Name and explain the sources of historical information.
   (b) Name each of the Oriental nations in the order of their development. State what was done by each for civilization.
2. (a) What were the causes of Greek colonization? What relation did the Greek colony have to the mother city?
   (b) Locate the chief centers of colonization and state for what each was famous.
3. (a) Trace out the history of the ancient Hebrew people and explain their service to humanity.
   (b) Describe the government and customs of the Spartans.
4. (a) Explain the origin, growth, and effect of the Delian Confederacy on Greek history. Show how it changed into the Athenian Empire.
   (b) Describe Athens in the time of Pericles.
   (c) Describe the intellectual greatness of Athens in the time of Pericles.
5. (a) Identify the following men and account for their greatness.
      1. Alexander
      2. Themistocles
      3. Aristides
      4. Plato
      5. Socrates.
   (b) Give an account of the invasion of Greece by Xerxes during the Persian Wars. Name important battles and tell in detail about one.
6. (a) What did Philip of Macedon accomplish for Macedonia?
   (b) Trace the march of Alexander against the Persians.
7. (a) Why was Europe better fitted than Asia to develop the highest civilization?
   (b) What mountain systems of Europe are not off-shoots from the central mass of the Alps?

GRAMMAR

(Eighth grade)

Answer ten questions.

1. Of what value is a good vocabulary?
2. How can a person acquire a good command of words?
3. What is an antonym? Write five words and give their antonyms.
4. Define synonym. Write five words and give their synonyms.
5. Define verb. A verb phrase.
6. Write a stanza of the Star Spangled Banner. Punctuate correctly and underline the verbs and verb phrases.
7. Define transitive verb. A direct object. Write a sentence containing a transitive verb and a direct object.
8. Name the eight parts of speech.
9. A boy is flying a kite.
   A crow is flying over the cornfield.
   Are the verbs in the above sentences transitive or intransitive? Why?
10. Use the verbs, lie and lay, sit and set, correctly in sentences.
11. Classify the nouns and verbs as to number in the following:
    - a. They came
    - b. We come
    - c. We have come
    - d. We had come
    - e. The house will be built.
    - f. The horses were running
    - g. They have been seen
    - h. We saw
12. What is a participle? Underline and give the grammatical construction of the participle in the following sentence:
    Crossing the street, I lost my hat.
13. What are the distinguishing marks of a verbal noun?
14. Give the construction of the verbal nouns in the following sentences:
    - a. To obey is a cardinal virtue.
    - b. Most boys like to play basketball.
    - c. Playing baseball is hard work.
    - d. I enjoy hearing pupils read.
15. Name and define the tenses. Tabulate the tenses of the verb "call."

ENGLISH

(High school, 1st year)

1. (a) In comparison with other languages is English old or new?
   (b) Why is the study of Latin important to us?
   (c) Give an example of a word derived from Latin and explain its parts.
   (d) Explain by illustration the use of a *prefix*.
   (e) Define literally: irregular—international.

2. (a) Explain the value of good pronunciation.
   (b) List 5 words that you have been mispronouncing.
   (c) Punctuation aids one in what way?
   (d) Illustrate the use of *one* punctuation mark.
   (e) State and illustrate the simple rule of spelling in regard to *ei* and *ie*.

3. (a) Name the 8 parts of speech. Illustrate.
   (b) Why should one make a good choice of words in writing or speaking?
   (c) Write an exclamatory sentence.
   (d) Give the plurals of: *datum*, *radius*, *lady*, *alumna*, *monkey*, *index*, *oasis*, *cargo*, *solo*, *volcano*.
   (e) Give feminine of: *abbot*, *hero*, *wizard*, *sir*, *lord*.

4. (a) How would you distinguish poetry from prose?
   (b) Give an example of *rhyme*.
   (c) Name 3 poetic qualities of Ancient Mariner? Illustrate.
   (d) Quote your favorite stanza from the poem, Ancient Mariner.
   (e) Name two descriptive passages from Vision of Sir Launfal.

5. (a) What is the purpose of the drama?
   (b) Name and give the dates of our greatest English dramatist.
   (c) Is "The Merchant" tragedy or comedy? Why?
   (d) Name the 4 stories which make up the plot.
   (c) Quote 3 passages from the play.
   (d) What is the climax of the play? Explain why.
   (e) Give setting of the story. Its source.

6. (a) Describe the theatre of Shakespeare's time, or characterize Shylock carefully, illustrating your points.

ENGLISH

(High school)

Answer ten questions.

1. (a) Discuss the work of three colonial prose writers.
   (b) Discuss the work of two colonial poets.

2. (a) What were the general tendencies of the literature of the Revolution?
   (b) Discuss the works of two writers who were closely connected with governmental affairs.

3. Who was the first American novelist? Tell about his works and characteristics as a writer.

4. (a) Give five important facts concerning the life of Irving.
   (b) Write in outline form a classification of Irving's works.

5. Give a detailed account of the life of your favorite American poet.

6. (a) Name three striking characteristics of the poetry of each of the six great American poets.
   (b) Name two of the best poems of each.

7. (a) Give five facts concerning the life of Poe.
   (b) Give ten striking characteristics of his work.

8. (a) Discuss the prose of Emerson, Lowell, and Holmes.
   (b) Quote two epigrams from Emerson's essays.

9. Write a paragraph on the subject—Thoreau's Individualism.

10. For what were the following noted: Walt Whitman, John Motley, Joel Barlow, Timothy Dwight, Francis Parkman, Thomas Bailey Aldrich, William Dean Howells, Bret Harte, Bayard Taylor, Sidney Lanier.

11. Name the authors of the following: Tales of a Wayside Inn, Commemoration Ode, The Prairie, The Prince of Parthia, Laus Deo, My Study Windows, The Last Leaf, Tampa Robins, The Blithedale Romance, Early Spring.

CIVICS

(Eighth grade)

Answer ten questions.

1. How does the Child Labor law govern the employment of children in Illinois?
2. What provisions are found in the U. S. Constitution in regard to the right to vote?
3. What are the voting qualifications in Illinois?
4. Explain how the President of the U. S. is elected.
5. How does a postman secure his position? What are some of the necessary qualifications?
6. Write the Preamble to the Constitution of the U. S.
7. In what particulars were the Articles of Confederation faulty?
8. When did the Constitution of the U. S. go into operation?
9. What is the purpose of a writ of Habeas Corpus?
10. How may a bill be passed over the President's veto?
11. State the duties of the County sheriff; the County Superintendent of Schools.
12. How many directors are there in school districts of less than 1,000 inhabitants?
13. What constitutes the Illinois Teachers' Examining Board? What are its duties?
14. State briefly the duties of the County Clerk.
15. What is minority or proportional representation? How is it used in Illinois?

CIVICS

(Eighth grade)

Answer ten questions:

1. What is a democracy? Name two. Compare our government to a ball team; explain an aristocracy through a ball team.
2. What does majority rule mean? Was it right for us to resist Britain in 1775? Why? Is a revolution ever dangerous?
3. Name five rights of American citizens. Name five duties of American citizens.

4. Where did we get our ideas of liberty? What was the Magna Charta?
5. Explain home rule in the United States. Who was responsible for the good or bad government?
6. Name the three branches of our government, and the representative of each.
7. Who may become president? What great law tells us this? Who is commander-in-chief of the army and navy?
8. When and where did the Constitution Convention meet? Who made the Constitution a power?
9. Write the Preamble to the Constitution.
10. Name five of the president's secretaries and tell who fills the offices.
11. Who makes treaties and issues passports? Who has charge of the mints? What is the difference between civil and political rights? When do our political rights begin?

CIVICS

(High school, 1st year)

1. What is an "unwritten constitution?" Give examples to show that we have one.
2. Name our colonial possessions and tell how each is governed.
3. Give qualifications and length of term for the President, a Senator, and a Representative.
4. What is meant by gerrymander, pocket veto, quorum pacifist, recall, neutral, arbitration?
5. Trace a bill through the process of becoming a law.
6. Of what does the Supreme Court consist? and what are its duties?

CIVICS

(High school)

Answer eight questions.

1. What differences did the framers of the constitution intend to create in the two houses?
2. What dangers are inherent in popular government?
3. "The federal government is one of limited power but within its own field it is supreme." Discuss.
4. What are the functions of the grand jury? Of the petit jury? What is a "hung" jury?
5. Discuss the origin of political parties in the United States. State the forces at work and the political leaders. Show how the different political parties, when in power, affect commerce and general prosperity of the nation.
6. Should we abandon our present electoral system? Give your reasons.
7. What is the difference between obeying a law and obeying a person?

8. Discuss metallic and paper money in the United States, stating the backing of each. Explain the European money market today on the basis of the above explanation.

9. What is the work of the National Committee?

10. What has the Washington Conference really accomplished?

GEOGRAPHY

(Seventh grade)

1. (a) Describe the formation of our continent.
   (b) Name the two great mountain systems and their smaller groups.

2. (a) Describe the size, shape and position of North America.
   (b) What was the extent of the Great Ice Sheet?

3. (a) Write an interesting paragraph about the Eskimos.
   (b) Name the New England States and give their capitals.

4. Describe the surface features, climate, rainfall and products of the Middle Atlantic States.

5. (a) Describe the mining of coal in the Middle Atlantic States.
   (b) Name the by-products of petroleum.

6. Name and describe the three chief industries of the Southern States.

PHYSICS

(High school)

Answer ten questions.

1. What is light? Cause of eclipse of moon? Show by drawing. Draw eclipse of sun.

2. Describe the rainbow. Show the formation in drawing. Show by drawing why the bow is curved.

3. What is heat? Temperature? Why does the boiling point vary?

4. What is the heat of fusion? Vaporization? How do these facts affect our life?

5. Describe a heating plant. Draw. (Either steam, water or air).

6. What is Probable Nature of Electrification? Theory of Lyden jar?

7. Describe a good open circuit battery and a good closed circuit battery.

8. Describe what you saw in X-ray.

9. How does a motor work? Describe the arc light. Incandescent.

10. Explain induction coil. Give practical uses of the coil. Give uses of transformer.

11. Describe either phone or telegraph in full.

12. Write 8 points either in favor or against the study of physics in high school.

ZOOLOGY

(High school)

Answer ten questions.

1. Name at least eight branches of the animal kingdom and an example of each class.
2. Discuss the classification of animals as to method used and basis for.
3. Explain how the amphibia stand between the fishes and the reptiles.
4. Give four illustrations showing how insects are adapted to their environment.
5. Why are the porifera a step higher than the protozoa?
6. Name an animal possessing one of the following:
    1. Alternation of generation.
    2. Bilateral symmetry.
    3. Complete metamorphoses
    4. Budding.
    5. Regeneration of lost parts.
7. Give an example to illustrate the struggle for existence and tell how the law of the survival of the fittest came to be established.
8. Why are the primates of such great importance?
9. What animal would you prefer to watch and study? Why?
10. Explain how Zoology helps you to realize the following objectives: 1. Health. 2. Vocation. 3. Use of leisure time.
11. What is your opinion as to the Theory of Evolution?

ARITHMETIC

(Eight grade)

Answer ten questions.

1. How much must I pay for U. S. 4¼% Liberty Bonds at 92, brokerage 1⅛, in order to have an annual income of $600?
2. How many blotters 6 inches long and 3½ inches wide can be cut without waste from a sheet of blotting paper 2 feet long and 14 inches wide?
3. What mathematical facts do the following numbers represent: 231, 7.92, 7000, 5280, 360, 62½, 31½, 32, 60, 16.
4. A baseball diamond, or infield, of regulation size for men is 90 ft. square. How long is a straight throw from first base to third?
5. Extract the cube root of 5832 and 148877.
6. Draw a figure to represent a section of land. (a) Number correctly the sections. (b) In a smaller drawing show the N. E. ¼ of the S. E. ¼ of Sec. 16.
7. Define mensuration, plane surface, rectangle, trapezoid, parallelogram.
8. How find the area of a parallelogram? How many square feet in a building lot 125 feet long and 50 ft. wide?
9. How many acres of land in a road 10 mi. long and 4 rods wide? What is the land in this road worth if land sells at $300 per acre?

10. Define circle, diameter, radius, circumference.

12. A boy measured the distance around a tree and found it to be 6¼ ft. How thick is the tree, correct to the nearest inch, where he made the measurement?

13. What is the lateral surface of a cylinder 15 inches high and 10 inches in diameter? What is its entire surface?

14. A silo (cylindrical) is 12 ft. in diameter and is filled to a depth of 18 ft. How many cubic feet of silage does it contain?

15. Define sphere, area of sphere, volume of a sphere. Considering the earth a sphere whose radius is 4000 mi., find the area of the earth's surface. Its volume.

BULLETIN NO. 10

BUREAU OF EDUCATIONAL RESEARCH
COLLEGE OF EDUCATION

RELATION OF SIZE OF CLASS TO SCHOOL EFFICIENCY

by

BUREAU OF EDUCATIONAL RESEARCH

Prepared, in part, from a report by P. R. Stevenson,
former Assistant, Bureau of Educational Research.

PRICE 50 CENTS

PUBLISHED BY THE UNIVERSITY OF ILLINOIS, URBANA
1922

TABLE OF CONTENTS

PAGE
Preface 3
I. Introduction: Relation of Size of Class to School Efficiency 5
II. Existing Conditions in Regard to Class Size 10
III. Relation of Size of Class in Elementary School to School Efficiency 16
IV. Relation of Size of Class in High School to School Efficiency 24
V. Suggestions for Educational Experimentation 37

PREFACE

This investigation was initiated by Mr. B. R. Buckingham, formerly Director of the Bureau of Educational Research. It was planned and executed by him with the assistance of Mr. P. R. Stevenson, a full time assistant in the employ of the Bureau of Educational Research during the school year of 1920-21. The present Director of the Bureau of Educational Research had no connection with the study until late in the summer of 1920-21. The portion of this report which deals with "existing conditions in regard to class size" and "the relation of size of class in high schools to school efficiency" is based upon tabulations made by employees of the Bureau of Educational Research under Mr. Stevenson's direction and included by him in a report submitted to the present Director of the Bureau of Educational Research. The chapter devoted to the relation of the size of class in elementary schools to school efficiency is based upon tabulations made from the original data under the immediate direction of the present Director. The concept of the efficiency ratio and the use of this concept in the interpretation of the data are entirely the work of the present Director of the Bureau of Educational Research. The conclusions are also his own.

Certain limitations of the investigation, which the report discusses in detail, cause the results to have a limited practical significance, but it is thought that the publication is jusified for two reasons. In the first place, the concept of school efficiency and the analysis of the conditions which must be considered in any investigation relating to school efficiency should be helpful to future investigators not only of the question of class size but also of questions of other phases of school procedure. In the second place, the report emphasizes the need for careful planning which will result in the control of all factors involved in the teaching situation. There is also emphasis upon the need for securing normal conditions if the results are to be interpreted with reference to the modification of practise. Such analysis and careful thinking are not only important phases of educational research but they are the foundation upon which both the data collected and the statistical manipulation of them are based.

This investigation was made possible through the cooperation of Superintendent Peter A. Mortenson, of the Chicago Public Schools, and of the school officials in certain other Illinois cities. Not only did they cooperate by permitting the collection of the data, but they actually made substantial contributions to the project by furnishing the test materials. The teachers in the schools concerned made a substantial contribution by scoring the tests and reporting them in a convenient form to the Bureau of Educational Research. The writer is glad to acknowledge the indebtedness of the Bureau of Educational Research to all those who have contributed to the project.

Walter S. Monroe, *Director.*

May 26, 1922.

RELATION OF SIZE OF CLASS TO SCHOOL EFFICIENCY

CHAPTER I

INTRODUCTION: ANALYSIS OF PROBLEM

The problem. The problem of this investigation is to study the relation of the size of class to school efficiency, or what is the effect upon the efficiency of the school when the size of class is increased or decreased within certain limits.

Definition of terms: Class. In the high school, a class is defined as the number of pupils who are assigned to a single teacher for instruction during a single class period. In the elementary school, unless the instruction has been departmentalized, a class is the number of pupils assigned to a room over which a teacher has charge. For instructional purposes, a teacher in the elementary school may divide a class into two or three groups, but the total number of pupils receiving instruction from her is considered a class, as the term is used in this study.

School efficiency. Educators have borrowed the term "efficiency" from industry and business. In these fields, efficiency is expressed by a fraction whose maximum is 1.00. The numerator of this fraction, or "efficiency ratio," is the output, and the denominator is the input, or educational investment. In education, the output of a school system consists of the changes produced in the pupils, i. e., the controls of conduct that the school engenders. The educational output for a semester or a year is the total of all the changes that have been produced in the pupils during the period due to the influence of the school. The educational investment includes many factors, such as buildings, equipment, textbooks, teachers, supervision, and general administration. Although it is not imperative to do so, it is probably best to think of both the output and the investment as being expressed in terms of the average for one pupil.

We are accustomed to refer to the educational output as the achievements of the pupils. By means of educational tests and other instruments, we measure these achievements in terms of arbitrary units. Units of one type are implied in school marks. Other units are defined by educational tests. In order to calculate the numerical value of the efficiency ratio, it would be necessary to determine the social value of the output in terms of dollars and cents or in terms of some unit which might be made common to both numerator and denominator. Obviously, we are not prepared to do this. However, we may use the ratio as a definition of school efficiency and inquire into the probable nature of the changes produced in it by the variations of certain factors upon which the educational output and educational investment depend. In making inferences concerning the fluctuations in the value of the efficiency ratio, it is necessary to remember that the production of certain achievements may be of little value when viewed in relation to our educational objectives. The attainment of certain levels of achievement may represent an educational output of considerable value, but advancement to higher levels may produce only slight increases in the value of the total output in this field. For example, the attainment of certain levels of ability in spelling has a distinct and relatively large social value, but advancement beyond these levels is accompanied by rapidly diminishing increments of value. Therefore, one must avoid the assumption that fluctuations in achievements are to be interpreted as having proportional values in terms of social worth.

Factors which affect school efficiency. For a given educational investment per pupil, the value of the efficiency ratio is affected when changes are made in the educational output. Methods of instruction, the plan of school organization, or the procedure of supervision may be modified even when there is no change in the investment. When modifications in the methods of using the investment result in changes in the educational output, there are resulting changes in the value of the efficiency ratio. On the other hand, it is possible that material changes may be made in the educational investment which are not accompanied by corresponding changes in the educational output. When this happens, the value of the efficiency ratio is changed, even though the actual educational output has remained constant.

In many cases, modifications in the method of using the educational investment are accompanied by changes in the magnitude of the educational investment as well as by changes in the educational output. Hence, we may have fluctuations occurring in both the numerator and the denominator of the efficiency ratio. It is possible that these fluctuations may be connected in such a way that the value of the efficiency ratio remains constant, or it may be that its magnitude will vary. Because we are not able to calculate a numerical value of the efficiency ratio, a careful analysis is required to determine the probable changes in it when variations occur in the numerator and the denominator simultaneously.

The achievements of pupils are materially affected by their general intelligence or capacity to learn. Since individual pupils and also groups of pupils have been shown to exhibit marked individual differences when measured with respect to this trait, it is necessary to make due allowance for differences in general intelligence when comparing different school units with respect to efficiency. In case this is not done an error of interpretation will be made by attributing a higher degree of efficiency to those units which consist of pupils of superior general intelligence.

The effect of varying the size of class. The size of class is one item of the plan of the organization. From this point of view, it may be considered one of the methods of using the educational investment. Consequently, we may expect to find that changes in the size of class produce variations in the achievements of pupils. The size of the class is, also, one of the factors which determines the educational investment. In the elementary school, where a class means the number of pupils assigned to a teacher, the cost of instruction per pupil varies inversely with the size of class.[1] In a high school the size of class does not completely determine the number of student hours of instruction which a teacher gives, but it is a potent factor in this determination. In general, an increase in the size of class in the high school will tend to result in a marked decrease in the educational investment per pupil. Hence, in studying the effect of varying the size of class upon the efficiency of the school, it is necessary for us to inquire into the resulting changes in both the educational output and the educational investment. It is only when we have done this that we are in a position to make inferences concerning the effect of variations in the size of class upon the efficiency of the school.

---

[1] In making this statement, no account is taken of investments made in supervision, in instruction by special teachers, and in equipment.

The practical importance of a study of the relation of class size to school efficiency. During recent years, school administrators have faced the problem of providing instruction for a rapidly increasing enrollment and, at the same time, of meeting the demands from teachers for increased salaries. In meeting these two demands, there has been a tendency to increase the number of pupils instructed by a teacher in order to keep the total educational expenditure within the income of the school system. In the elementary school, and to a considerable extent in the high school, the number of pupils instructed by a teacher has been increased by increasing the size of classes. It is obvious that pupils in large classes have less opportunity for recitation and, in general, receive less individual attention from the teacher, both within and outside of the class period. Thus, the question has naturally been raised concerning the effect upon the efficiency of a school when the size of the class is increased. In the secondary school added emphasis is given to the question because certain accrediting agencies require that the size of the class not exceed a certain fixed maximum.

Connection between size of class and methods of instruction. It is necessary to bear in mind that, when the number of pupils instructed by a teacher is increased, there is a corresponding increase in the amount of work required of the teacher, unless there are compensating changes in the methods of instruction. For example, according to our present methods, it is customary to require a great deal of written work of pupils studying English composition. A teacher is expected to read with considerable care the compositions submitted by pupils and to provide a systematic procedure for correcting the errors made. Thus, an increase in the number of pupils to be taught increases the work required of the teacher unless the number of compositions required from each pupil is reduced or a different system of handling them is used. Much the same conditions prevail in a number of other subjects in which notebooks or other written work of some sort are customarily required. Similar statements can be made with reference to individual work with pupils outside of the regular class period. It is obvious that there is a limit to the amount of work which may legitimately be required of a teacher. When the optimal teaching load has been reached, any increase in the number of pupils assigned to a teacher should be expected to be accompanied by compensating changes in the procedure of instruction. When such changes are made in the teaching procedure, they, as well as the size of the class, must be considered in respect to the effect upon the efficiency of the school.

Limitations of the present investigation. In this investigation of the effect of varying the size of class upon school efficiency, it was intended that all other factors which affect either the achievements of pupils or the educational investment should be kept constant, i. e., they should be the same for classes of different sizes. No information is available to show the extent to which this intention was realized except in the case of the quality of the pupil material. Group intelligence tests were used to secure equivalence of capacity to learn in the two groups of pupils. The possibility of a lack of equivalence of such factors as home environment, nationality, attitude toward school work, previous school experience with respect to size of class, time of day (applies only to high school), etc., makes it necessary to exercise due caution in interpreting the results of the investigation.

So far as the present writer is aware, the instruction that prevailed in the groups concerned in this investigation involved no unusual features, and the same methods of instruction were followed in the two types of classes. It is possible that the highest degree of efficiency for classes of a given size would be attained if the methods of instruction were selected with particular reference to this size of class. It would not be surprising to find that methods of instruction which were most effective in small classes would be considerably less effective with large classes, and that methods well suited to the handling of large classes would not give the best results when used with small classes. This possibility was not considered in this investigation. Hence, the findings should not be accepted as final. There is still need for an investigation in which methods of instruction are adapted to size of class.

Finally, it must be remembered that the problem of the size of class is not entirely a problem of the efficiency of the school. It is also a problem of the teacher. It is not humane and it is not socially profitable to assign teaching loads so heavy that teachers become overworked. The problem of the teacher was not considered in this investigation.

CHAPTER II

EXISTING CONDITIONS IN REGARD TO CLASS SIZE

In order that the significance of the two investigations to be reported in the following chapters may be more fully appreciated, certain facts are presented, concerning the size of class under present school conditions in Illinois.

Size of class in elementary schools in Illinois outside of Chicago. Data with reference to the size of class in the elementary schools of the state, outside of Chicago, were secured by sending a questionnaire to the superintendent of public schools in all cities and towns listed in the Illinois School Directory for 1920-21 as having six or more elementary teachers. The questionnaire asked for the number of elementary teachers having classes of the following sizes: less than 20, 20 to 29, 30 to 39, 40 to 49, and 50 and over. This information was requested for each of the school years of 1918-19, 1919-20, and 1920-21. Complete reports were received from 180 cities and towns. These are summarized in Table I. The total number of classes for which a report was secured varied from 9,422, 1918-19, to 10,403, in 1920-21. The median size of class varied from 41.4 pupils, 1918-19, to 43.3 pupils, in 1919-20. In 1920-21, the size of class was slightly less than that for the preceding year. This table also shows that, in the cities reporting, slightly more than 3 percent of all classes contained 50 or more pupils. On the other hand, between 7 and 8 percent of the classes had less than 30 pupils. The greatest change in regard to the size of class during this period was in the marked decrease in the number of classes having between 30 and 39 pupils. During the school year, 1918-19, the size of slightly more than one-third of the classes fell within these limits. During the other two school years, covered by this study, less than one-fifth of the classes came within these limits of size. This change is offset by a corresponding increase in the number of classes having between 40 and 49 pupils. Although the span of years covered by this table is insufficient to justify conclusions with reference to the trend of the size of class, Table I suggests that there is a tendency, in

TABLE I. SIZE OF CLASSES IN THE ELEMENTARY SCHOOLS OF 180 ILLINOIS CITIES FOR 1918-21.

Number in Classes	1918–19		1919–20		1920–21	
	Number of Classes	Percent of Classes	Number of Classes	Percent of Classes	Number of Classes	Percent of Classes
50..........	313	3.3	342	3.5	350	3.4
40–49.......	5134	54.5	6865	70.0	7161	68.8
30–39.......	3264	34.6	1845	18.8	2050	19.7
20–29.......	598	6.3	646	6.6	719	6.9
Less than 20	113	1.3	110	1.1	123	1.2
Total.......	9422	100.0	9808	100.0	10403	100.00
Median.....		41.4		43.4		43.2

the elementary schools of Illinois outside of Chicago, to assign from 40 to 50 pupils to a teacher.

Size of class in Chicago public schools, October, 1920. Because the two investigations to be reported later were carried on, for the most part, in certain schools in Chicago, it was thought desirable to compile separately the facts relating to class size in Chicago. The information was taken from the records of the superintendent of schools. One-half of the elementary schools were selected at random, but all of the high schools were included. Information, with reference to the size of class in each grade, is summarized in Table II. Below the ninth grade, the classes are noticeably larger than in the high school. The median size of class is approximately 46 pupils. In the high school, a greater degree of variability is shown in the size of class. Approximately 14 per cent of the classes have fewer than 20 pupils. The median size of class for the high school is slightly over 30 pupils. It will be noted that the classes for the first year are larger than those for the following years. Of the 183 classes, 178 reported as having from 50 to 54 pupils, are classes in physical training. A number of these have more than 54 pupils.

The size of high school classes is given by subjects in Table III. The median size of class for the different subjects ranges from 20.0,

TABLE II. SIZE OF CLASSES IN CHICAGO SCHOOLS, OCTOBER, 1920.

Grades	Less than 10	Number in class									55 and over	Total	Median
		10–14	15–19	20–24	25–29	30–34	35–39	40–44	45–49	50–54			
1			3	1	6	20	40	120	248	87	36	561	46.8
2			1	2	1	6	11	102	247	38	1	409	46.7
3			1	1	1	3	18	96	211	48	3	382	46.7
4				1	2	3	15	90	207	39		257	46.6
5			1			2	4	79	215	48	1	350	47.1
6			3	3	1	2	12	103	183	27	4	338	46.2
7					2	2	15	92	180	32	3	326	46.4
8				1		4	54	79	117	15	48	318	44.8
Total Elementary Schools			9	9	13	42	169	761	1618	334	96	3041	46.6
9	36	108	224	556	613	850	683	254	98	92		3474	32.3
10	24	84	202	342	378	401	283	136	17	69		1936	29.3
11	14	60	90	178	148	189	119	31	10	19		858	27.8
12	15	27	64	67	81	93	77	29	2	3		458	28.4
Total H. S.	89	279	580	1143	1220	1533	1162	450	87	183		6726	30.2

for German, up to 34.5, for arithmetic. If physical education is included, the maximum median class is 43.8. However, the most significant aspect of the table is the wide variation in the size of class for a given subject. With few exceptions, there are, in each subject, classes having 10 or less pupils and also classes having more than 45 pupils. The exceptions are shop work, in which the number of pupils is probably limited by the equipment, office practise, of which there are only 37 classes in the entire school system, chemistry, home economics, botany, zoology and agriculture, in which equipment again probably limits the size of class. In German there are no classes having more than 34 pupils, but there are only 8 classes in the entire system.

Opinions of city superintendents in regard to the best size of class. A questionnaire was sent to the city superintendents of public schools in all cities in the United States having a population

TABLE III. SIZE OF CLASSES BY SUBJECTS IN CHICAGO HIGH SCHOOLS, OCTOBER, 1920.

Subject	Less than 10	Number of pupils in class								Total	Med-ian
		10–14	15–19	20–24	25–29	30–34	35–39	40–44	45–49	50–54		
Botany, Zoology and Agriculture	6	12	24	46	49	51	22	5			215	27.0
French	6	14	33	51	35	38	15	8	2	1	203	24.7
Home Economics	7	27	42	57	54	18	11	3			219	22.9
Geography	1		17	17	19	14	19	8	1		96	28.4
Typewriting	1	1	12	36	50	70	74	29	11		284	33.0
Art	7	22	35	44	44	59	28	11	2		252	27.0
Physiology			1	13	17	36	44	12	1	1	125	34.4
German	2	1	1		3	1					8	20.0
Chemistry	2	5	15	21	19	39	3	1			105	27.5
Office Practise	3	4	2	11	4	6	7				37	24.3
Spanish	2	15	18	28	66	59	41	16	2		247	29.6
Physical Education	5	3	19	21	42	61	40	47	39	178	455	43.8
Latin	3	15	20	31	46	52	56	21	1		245	30.8
Mechanical drawing	22	29	63	136	97	67	31	9	1		455	24.2
Algebra	1	2	19	36	73	133	115	27	2		408	32.7
Geometry	1	9	16	36	62	79	51	13	3		270	30.6
Arithmetic			5	18	25	43	57	24	1		173	34.5
Bookkeeping	2	15	14	23	28	27	30	17	1		157	29.4
Stenography	1	9	23	55	53	72	65	17	4		299	30.6
Physics	1	2	12	29	18	40	22	5	1		130	30.4
Shop	8	66	94	228	101	16	3				516	22.0
History	4	13	27	49	37	92	86	63	3	1	375	33.1
English	3	15	60	143	228	352	278	85	10	2	1176	32.0
General Science	1		8	14	50	108	65	28	2		276	33.0

TABLE IV. IDEAL SIZE OF CLASSES AS INDICATED BY 270 SUPERINTENDENTS IN CITIES OF 25,000 OR MORE POPULATION.

Score	Grades			
	1, 2, 3,	4, 5, 6,	7, 8, 9,	10, 11, 12
55–59	1			
50–54		1	1	
45–49	2	2		1
40–44	16	18	15	2
35–39	53	73	35	4
30–34	95	95	104	33
25–29	52	42	58	95
20–24	20	21	48	95
0–19	30	18	9	16
Total......	269	270	270	246
Median....	31.7	32.8	31.0	25.6

of 25,000 or more, as shown by the directory issued by the Bureau of Education. The questionnaire asked the superintendents to indicate the ideal size of class in their opinion for grades one to three, four to six, seven to nine, and ten to twelve. Replies from 270 cities are summarized in Table IV. One of the most interesting things about this table is the wide range of opinion which it indicates. A considerable number of superintendents would have fewer than 20 pupils in each class. Other superintendents appear to consider classes of 40 or more ideal. One superintendent indicates that he would be satisfied with classes of 55 pupils in the primary grades. The ideal median size of class below the high school ranges from 31.0, for the junior high school, to 32.8, for the intermediate grades. In the senior high school, the ideal median of class is 25.6 pupils.

The problem of class size. Although Table IV is based upon replies from superintendents distributed over the United States, and the preceding tables refer to existing conditions in Illinois, we are probably justified in pointing out that a marked difference exists between theory and practise. The prevailing practise in the elementary schools of the state outside of Chicago centers around classes having from 40 to 45 pupils. The median ideal size of class is only slightly above 30. Thus, it appears that the practical problem in which school superintendents are interested relates to a determination of the relative efficiency of classes enrolling from 25 to 35 pupils as compared with classes enrolling from 40 to 45 pupils.

For the high school, we have no data relating to the size of class in Illinois except for the city of Chicago. The median ideal size is approximately 25. The median actual size is approximately 30. It, therefore, appears that the practical problem in the high school relates to the relative efficiency of classes enrolling from 20 to 30 pupils as compared to those enrolling from 25 to 35.

CHAPTER III

RELATION OF SIZE OF CLASS IN ELEMENTARY SCHOOL TO SCHOOL EFFICIENCY

General plan of the study in the elementary school. In order to study the effect of variations in the size of class upon the achievements of pupils, it is necessary to hold constant or to measure the other factors which affect their achievements. It was planned to keep the teacher constant by having both a large class and a small class taught by the same teacher. Since the plan of organization in the elementary school makes it impossible for the same teacher to instruct two classes at the same time, it was necessary to have a teacher instruct the two types of classes during two consecutive semesters. It was arranged to have some of the teachers instruct a large class during the first semester and a small class during the second semester. Other teachers instructed a small class during the first semester and a large class during the second semester.[1]

In order to keep the pupil material as nearly constant as possible, "one hundred percent promotion" was secured at the end of the first semester in all of the experimental groups. When a teacher instructed a large class during the first semester a number of pupils were sent to another teacher at the beginning of the second semester. The pupils remaining formed a small class. In doing this, an effort was made to select pupils so that those remaining would form a small class having approximately the same average mental age and the same variability of this trait. When a teacher instructed a small class during the first semester, pupils were added at the beginning of the second semester, but care was exercised to have these pupils such that the average mental age of the class would not be materially affected.

This investigation, which began in October, 1920, was confined to classes in grades II, V, and VII. Some of the experimental groups were organized in the B sections of these grades and the others in the A sections. At the beginning of the second semester the B sections became A sections of the same grade and the A sections became B sections of the next higher grade. We shall, however, refer to the grades simply as II, V, and VII. Data for only those pupils who attended both the large class and the small class and who took all of the tests given are included in the following tabulations.

---

[1] This investigation was carried on in five elementary schools in Chicago: Washington, Cleveland, Lowell, Farragut, and Hibbard.

The size of the experimental classes. In the second grade there were eleven experimental classes, three small the first semester and large the second, and eight of the opposite type. If one class, enrolling only 18 pupils when considered a large class, is excluded, the small classes range from 33 to 44 and the large from 45 to 54. The differences in the size of the paired groups range from 4 to 13, the average being approximately eleven pupils. In the fifth grade there were thirteen experimental classes, three small the first semester and large the second, and ten of the opposite type. The small classes range in size from 33 to 45 and the large from 42 to 52. The differences in the size of the paired groups range from 4 to 14, the average difference being approximately 9. In the seventh grade there were only five experimental classes, three of one type and two of the other. The small classes ranged from 35 to 44 and the large from 42 to 49. The average difference in size was approximately 7. It should be noted that in both the fifth and seventh grades there is some overlapping in the size of the two types of classes. Some "large classes" are smaller than certain "small classes."

Data collected. In the second grade, the Dearborn Group Intelligence Test and Pressey Primer Scale were given at the beginning of the experiment. In this grade, achievement was measured by giving the Indiana Scale of Attainment, No. 1. Form 1 was given in October, Form 2 in January, and Form 1 was used again at the end of the year. In grades V and VII, the Illinois General Intelligence Scale was used. The achievements of the pupils were measured in arithmetic, silent reading, language, and spelling. In arithmetic and reading, the measurements were secured by means of the tests included in the Illinois Examination. Form 1 was used for the first and third testings and Form 2 for the second. In language, Charters' Diagnostic Language Test for pronouns was used. Form 1 was given in October and in May. Form 2 was used for the January testing. In spelling, 20 words were selected from columns N and R of the Buckingham Extension of the Ayres' Spelling Scale.

For the first and third testings, the words were selected by beginning at the bottom of these columns and choosing alternate words. The words for the second testing were taken from these columns, beginning with the next to the last word and taking alternate ones.

Administration of tests and collection of data. All the tests were administered and scored by the teacher. As a preparation for this work, the teachers were called together and given definite instructions concerning the nature of the tests and the plan of administration. In this connection, the tests were administered to the teachers in order to illustrate to them the procedure to be used with the pupils. All of the tests are highly objective with reference to the scoring, and samplings of the test papers failed to reveal any large errors in this work. The teachers reported the data for each pupil on an individual record card. This card contained spaces for each score and each test as well as for data with reference to the size of class in which the pupil was taught during each semester. The dates of testing were approximately as follow: October 20th, February 20th, and May 20th.

Method of summarizing the data. The data summarized were limited to the scores of only those pupils who were members of both a large class and a small class, and who were present at all three testing periods. The scores of all pupils in a grade (including both A and B sections), who had been taught in the same type of class, were assembled for each of the three testings. For example, the October arithmetic scores for all fifth grade pupils who were taught in large classes during the first semester and in small classes during the second semester were assembled in one distribution. Another distribution was made for the January scores and a third one for the May scores. Three corresponding distributions were made for the arithmetic scores of the pupils who were taught in small classes during the first semester and in large classes during the second. Thus, there were obtained six distributions for each achievement score. One measure of the gain in achievement made by a group of pupils during the first semester was found by subtracting the *average* of the October scores from the *average* of the January scores. Another measure of the gain was found by subtracting the *median* October score from the *median* January score. In a similar manner, the gains for the second semester were obtained by subtracting the average and the median January scores from the corresponding May scores.

In calculating these gains, no account was taken of the possible non-equivalence of the different forms of the tests used. In fact, no accurate information concerning the equivalence of the duplicate forms is available, except for the tests in reading and arithmetic. The duplicate forms of these two tests have been shown to be approximately equal.[2] Since Form 1 was used twice, and the average and the median scores calculated from it were used both as subtrahends and minuends, any non-equivalence of the forms will not affect the comparisons of gains made in the following table.

The scores of the different tests are expressed in terms of different units. Thus, before any combination of the results from the different tests can be made, it is necessary to express the gains in terms of a common unit. The usual assumption in such cases is that the standard deviation of the distribution of scores represents the same increment of ability for one test and in one grade as in another. On the basis of this assumption, the stardard deviation was calculated for six of the different distributions of scores for each test in a given grade, and the average of these six standard deviations was used as a divisor to reduce the gains to a basis of a common unit. For example, during the first semester the fifth grade pupils taught in large classes in arithmetic made a gain of 12.0 points.[3] During the second semester they made a gain of 7.2 points. The gains for the pupils taught in small classes in arithmetic were 6.5 and 5.65. The average standard deviation of the six distributions of arithmetic scores is 17.441. Dividing these gains by this average standard deviation, we secure as quotients the entries (.68, .41, .36, and .32) to be found in Table V.

In calculating the average gains for the two types of classes, the simple average of the two gains has been used rather than the weighted average, although the two groups are not even approximately equivalent in size. Since our purpose in taking this average is to eliminate any differences in the course of study or in the edu-

---

[2] Monroe, Walter S. "The Illinois Examination." University of Illinois Bulletin, Vol. 19, No. 9, Bureau of Educational Research Bulletin, No. 6. Urbana: University of Illinois, 1921.

[3] These gains were calculated from the average of the scores.

TABLE V. THE GAINS MADE BY PUPILS IN LARGE CLASSES AND IN SMALL CLASSES. CLASSES OF ALL SIZES INCLUDED

	No. of pupils	Average size of class		Arithmetic Gains		Dif.	Reading Comprehension Gains		Dif.	Reading Rate Gains		Dif.	Language Gains		Dif.	Spelling Gains		Dif.	Totals		Dif.
		Large	Small	Large	Small		Large	Small		Large	Small		Large	Small		Large	Small		Large	Small	
GRADE II.																					
Average Scores:																					
Large 1st S.	244	44.8	33.1	1.89	.59		1.17	.59		1.27	.38					1.65	.88		5.98	2.44	
Small 1st S.	109	48.0	38.0	.30	1.48		.53	1.09		.51	1.00					−.13	1.83		1.21	5.40	
Average				1.09	1.04	.05	.85	.84	.01	.89	.69	.20				.76	1.36	−.60	3.59	3.93	−.34
Median Scores:																					
Large 1st S.	244	44.8	33.1	2.25	.70		1.28	.80		1.09	.60					2.06	.95		6.68	3.05	
Small 1st S.	109	48.0	38.0	.33	1.94		.80	1.05		.53	.97					.50	1.64		2.16	5.60	
Average				1.29	1.32	−.03	1.04	.93	.11	.81	.78	.03				1.28	1.29	−.01	4.42	4.32	.10
GRADE V																					
Average Scores:																					
Large 1st S.	325	45.8	35.7	.68	.41		.20	.88		.09	.72		.89	1.22		.02	.62		1.88	3.85	
Small 1st S.	132	51.0	44.0	.36	.32		.55	.38		.76	.28		.93	.96		.38	.29		2.98	2.23	
Average				.52	.37	.15	.38	.63	−.25	.43	.50	−.07	.91	1.08	−.17	.20	.46	−.26	2.44	3.04	−.60
Median Scores:																					
Large 1st S.	325	45.8	35.7	.78	.36		.32	1.12		.08	1.14		.79	1.77		.01	.67		1.98	5.06	
Small 1st S.	132	51.0	44.0	.21	.43		.75	.29		.93	−.03		.98	.99		.43	.26		3.30	1.94	
Average				.50	.40	.10	.54	.71	−.17	.51	.56	−.05	.89	1.38	−.49	.22	.46	−.24	2.66	3.51	−.85
GRADE VII.																					
Average Scores:																					
Large 1st S.	104	45.0	36.0	.85	1.29		.69	.72		.74	.45		.57	1.26		.92	.21		3.77	3.93	
Small 1st S.	73	47.0	42.0	.53	.47		.42	.19		.64	−.61		.62	1.00		.15	.44		2.36	1.49	
Average				.69	.88	−.19	.56	.46	.10	.69	−.08	.77	.59	1.13	−.54	.54	.33	.21	3.07	2.72	.35
Median Scores:																					
Large 1st S.	104	45.0	36.0	.92	1.28		.76	.86		.67	1.01		.50	1.34		.98	.47		3.83	4.96	
Small 1st S.	73	47.0	42.0	.65	.66		.45	.21		.56	−.45		1.09	.99		.24	.64		2.99	2.05	
Average				.79	.97	−.18	.61	.54	.07	.62	.28	.34	.79	1.17	−.38	.61	.56	.05	3.42	3.52	−.10
Totals of averages of Average Scores				2.30	2.29	.01	1.79	.193	−.14	2.01	1.11	.90	1.50	2.21	−.71	1.50	2.15	−.65	9.10	9.69	−.59
Totals of averages of Median Scores				2.58	2.69	−.11	2.19	2.18	.01	1.94	1.62	.32	1.68	2.55	−.87	2.11	2.31	−.20	10.50	11.35	−.85

cational opportunities offered in the two semesters and also any practise effect due to acquaintance with the tests, it seemed unwise to weight the averages on the basis of the number of pupils in the two groups. To have used the weighted averages in this case would have resulted in giving greater weight to the gains made by one group of pupils simply because this group happened to be larger.

Achievements of the two groups approximately equal. Table V[4] summarizes the data with reference to the gains made by the two groups. The column headed "Number of pupils" gives the number of pupils whose records were used in the tabulations. (A pupil's record was discarded if he was not a member of both the small class and the large class and if he did not take all tests.) The average size of class is computed from the total enrollment. The computation of the gains has just been explained. In interpreting the table, attention should be focused upon the differences. A positive difference means that the large class is superior in achievement, and a negative difference that the small class is superior. At the bottom of each difference column the differences, calculated from the averages and also from the medians, are summarized. This summary is a total and not an average. To find the average it is necessary to divide by 3. All of the totals of the differences fall between +1.00 and —1.00. Six of the 10 differences are negative, and only in the case of reading rate are the difference between the averages and the difference between the medians both positive. In the last three columns of the table, we have the totals and not averages. To find the average, it would be necessary to divide by 4 in the second grade and by 5 in each of the other two grades. Here, again, the negative differences predominate, although none of them are very large. The last two entries in the last column are essentially grand totals and may be considered to summarize the entire table. To find the average difference, each of these numbers should be divided by 14. The quotients obtained would be —.04 and —.06. Thus, in general, this table indicates that there is little if any superiority in the achievements of pupils in the small classes over those of pupils in the larger classes.

---

[4] The entries in the column headed "Reading Rate" in the second grade are based upon the Pressey Word Recognition Test.

An examination of Table V reveals the fact that, in general, the gains between the first and second testings are much larger than those between the second and third testings. This condition emphasizes the necessity for equalizing the effect of acquaintance with the test and practise effect. If all of the experimental groups had been taught as large classes the first semester and small classes the second, the gains for the large classes would greatly exceed the gains for the small classes; but this would be due primarily to the effect of acquaintance with the test and the practise effect.

When Table V is examined with reference to the conditions in the different grades we find that the gains are relatively greater for the small classes in the fifth grade than in either the second or the seventh grade. However, the number of pupils is so small for the two groups in the seventh grade and the groups differ so little in size that only slight significance can be attached to the results. Even in the second and fifth grades it is unfortunate that the experimental groups are not more nearly equal in size. It is possible that if the experiment had included a larger number of classes which were small the first semester and large the second different results might have been obtained.

When the gains for the different subjects are examined we find that in language the gains for the small classes are consistently greater than the gains for the large classes. This is not true for any other subject. Although in spelling the total of the gains is distinctly negative, in both arithmetic and reading comprehension the total when computed by one method is negative and in the other case is approximately 0. Reading rate in the seventh grade is the only case in which the large class is distinctly superior in achievement.

Conclusion: the relation of the size of class to school efficiency. Since Table V indicates that, on the whole, there is little difference between the achievements of the pupils when taught in large classes and their achievements when taught in small classes, one might infer that the efficiency of a school would be materially increased by the formation of large classes, because the educational output would be approximately the same and the educational investment would be materially decreased. However, it is doubtful that the present investigation justifies such a conclusion. In the first place, it is obvious that only certain achievements of pupils have been measured. Even in the fields of the four subjects in which tests were given we are not justified in claiming that all achievements of the pupils were measured. The arithmetic tests used were confined to the operations and to only certain types of examples within this division of arithmetic. In silent reading, the test used is very limited in scope. Similarly, the tests in language and spelling possess very definite limitations with respect to scope. There is some justification for assuming that the measurements made may be considered indices of the total achievements of the pupils not only in the fields of the four subjects in which the tests were given but also in the field of instruction in the grades concerned. However, the thesis that the measures of achievement secured in this investigation are indices of the total achievement is largely an assumption, and in interpreting the results it is necessary to recognize this fact. It is possible that, if other tests had been used or if the achievements of the pupils had been more completely measured by including tests in other subjects, the results might have been different.

In the second place, it must be remembered that the size of the "small classes" was not less than 33 (with one exception), and in a few cases the enrollment was as much as 44 or 45. The large classes ranged in size from 42 to 54. The average difference between the pairs of experimental groups ranged from 7 in the seventh grade to 11 in the second grade. These conditions with reference to the size of the experimental groups constitute a very significant limitation of the investigation. One is not warranted in making inferences from the facts of Table V with reference to the relative efficiency of classes of 20 to 25 pupils as compared with classes of 35 to 45 pupils. No application should be made except within the limits of size defined by the experimental groups.

CHAPTER IV

RELATION OF SIZE OF CLASS IN HIGH SCHOOL TO SCHOOL EFFICIENCY

General plan of the study in the high school. In the investigation in the high school it was arranged to have both a large class and a small class in a given subject instructed by the same teacher during the same semester. This made it necessary to restrict the experiment to teachers who were instructing two or more sections of the same subject. When a teacher was instructing two sections, pupils were shifted on the basis of their intelligence scores so that the average quality of pupil material was approximately the same in the two sections. Thus, both classes would include some bright, some medium, and some dull pupils. When a teacher had four sections of the same subject, the pupils were shifted so that a large class and a small class would be obtained, consisting of relatively bright pupils. The less capable pupils were also divided into two classes, one large and one small.

In establishing the two types of class, there was considerable variation in the size of both the large classes and the small classes. The small classes varied in size from 12 to 26 pupils. The large classes varied in size from 23 to 45 pupils. The average size of the large classes was 36.5 and that of the small classes, 20.8. The differences in the size of the paired classes ranged from 6 to 26.

Source of data. This study was carried on in four large high schools in Chicago and in the high schools of three other Illinois cities.[1] During the first semester of 1920-21, the experiment was carried on in beginning tenth grade classes. During the second semester of that year the study was confined to classes completing the first year of high school work. In the following tables no distinction is made between classes for the two grades. Records were secured for 67 pairs of classes, enrolling 3,821 pupils. The project was begun by giving the Terman Group Test of Mental Ability to all pupils concerned, about October 15, 1920. Some pupils in the high schools outside of Chicago were given the Illinois General Intelligence Scale or the Chicago Group Intelligence Test. As soon as the results of the testing could be assembled, the large classes and small classes were arbitrarily formed, following the method indicated above.

---

[1] The high schools in Chicago were Lane Technical, Tilden Technical, Harrison Technical, and Hyde Park. The three other Illinois cities were Macomb, Shelbyville, and West Aurora.

The educational output, or the achievements of the pupils, was measured by requiring each teacher to give the same final examination to both types of classes. A check upon this measurement of achievement was secured by using the "term grades" of the pupils. It is generally recognized that "term grades," as well as examinations set by teachers, are highly subjective. However, in this case the same teacher administered the same examination to both the small class and the large class. The same teacher also gave the "term grades." Thus, there is in no place a comparison between either "term grades" or "examination grades" given by different teachers. This tends to eliminate the subjective factor of these measures. In addition, it may be noted that we are concerned with the average "grades" of relatively large groups of pupils and not with the "grades" of individual pupils.

A limitation. The plan of carrying on the experiment implies the assumption that the achievements of the pupils in the large classes were equal to the achievements of the pupils in the small classes at the beginning of the experimental period. There was no attempt to measure the achievements of pupils in the subjects concerned at the beginning of the experiment. In the first semester, the two types of classes were not organized until after the high schools had been in session several weeks. In the case of the classes used during the second semester, the pupils had received an entire semester of instruction in regular classes. It is true that the sections were formed so that the average general intelligence of the paired classes was approximately equal, but this probably does not justify the assumption of equivalent achievements.

Details of administration. The intelligence tests were administered and scored by the teachers. As preparation for this work, the teachers were called together and given definite instructions concerning the nature of the tests. The tests were also administered to them. They were then required to score their own papers. All the tests used were highly objective with reference to scoring, and a sampling of the test papers of the pupils failed to reveal any large errors in this work. Since, in the use of the intelligence test scores, comparisons are always made between the pupils or groups of pupils under the same teacher, variations in the administration of the tests, due to differences between teachers, would not be significant.

At the close of each semester the teachers were asked to report both the "final grade" and the "examination grade" for each student.[2] In most cases, the "examination grades" were reported in terms of percents. The "final grades" were generally reported in terms of letters, as follow:

S=Superior
E=Excellent
G=Good
F=Fair
D=Failure

For the purpose of combining "grades," these letters were assumed to represent the following numerical ranges on a scale of 100 percent:

S is equivalent to 95 and over
E is equivalent to 85 to 94
G is equivalent to 75 to 84
F is equivalent to 65 to 74
D is equivalent to 55 to 64[3]

Plan of summarizing data. The data collected were summarized to show the differences, if any, which existed between the final achievements of pupils in the large classes and of pupils in the small classes. Two methods of summarizing were employed. In the first, the achievements of all pupils were considered. According to the second method, the records considered were limited to those of pupils in the large classes who could be paired with pupils having identical scores on the intelligence test in the corresponding small classes. For both of these methods two tabulations have been made. The first includes all classes, and the second only those pairs of classes in which the large class was at least twice the size of the small class.

---

[2] The following tables, in which the data for high school classes are summarized, indicate that the "examination grades" were not received from certain classes.

[3] The midpoints of these intervals were presumably used as the numerical equivalents of the grades expressed in terms of letter. However, Mr. Stevenson's report yields no information on this point.

Differences in achievements when all pupils are considered. Table VI illustrates the first method of summarizing the data for 22 pairs of English classes. In the second and third columns of this table the enrollment in the large classes and in the small classes is given. The quantities recorded in the three columns headed "Difference" are found by subtracting the quantities in the two columns immediately preceding the difference column. The number for the small class is, in every case, taken from that for the large. Therefore, a positive difference means that the large class is superior in the trait concerned, and a negative difference, that the small class is superior. The line at the bottom of the table gives the average for each column. These averages may be taken as summarizing the data collected from these 22 pairs of English classes, although the items combined are not entirely comparable. For example, different general intelligence tests were used in different classes. The present writer has not been able to ascertain the particular intelligence test given to any pair of these 22 pairs of classes. It is, however, difficult to explain the extreme differences between the average intelligence scores of classes 10 and 11 on any basis other than the use of different tests in these two pairs of classes. Furthermore, it is not unlikely that different passing marks are in use in the different schools in which the pairs of classes were taught. If this is the case, in the case of different pairs of classes, both the average term grades and the average examination grades are on different scales.

It should be noted that, although an effort was made to organize a large class and a small class so that the average intelligence scores would be approximately the same for the two classes, this was not always accomplished. Because of conflicts or other conditions that could not be disregarded, it was not always possible to shift pupils from one section to the other so as to set up the desired class organization.

Table VII summarizes the averages for the classes in the different subjects. In interpreting this table, it is necessary to bear in mind that, with the exception of English and algebra, the number of

TABLE VI. COMPARISON OF AVERAGE INTELLIGENCE SCORES, AVERAGE TERM GRADES, AND EXAMINATION GRADES OF 22 LARGE AND 22 SMALL ENGLISH CLASSES.

Teacher	Size		Average intelligence score		Difference	Average term grade		Difference	Average examination grade		Difference
	Large class	Small class	Large class	Small class		Large class	Small class		Large class	Small class	
1	36	12	58.7	57.2	1.5	85.8	77.5	8.3	85.3	77.5	7.8
2	37	17	95.2	99.7	−4.5	69.4	71.2	−1.8			
3	34	17	54.0	50.7	3.3	76.8	79.4	−2.6			
4	40	18	111.5	114.2	−2.7	82.1	81.2	0.9	77.8	85.1	−7.3
5	44	18	92.3	88.4	3.9	72.8	72.2	0.6	74.1	73.6	0.5
6	38	19	98.1	100.2	−2.1	74.2	80.6	−6.4	71.6	80.0	−8.4
7	35	19	138.7	137.1	1.6	77.1	81.6	−4.5	79.8	83.2	−3.4
8	32	19	94.9	87.8	7.1	72.4	75.6	−3.2	69.2	74.1	−4.9
9	41	20	77.4	72.7	4.7	75.8	77.0	−1.2	75.8	75.3	0.5
10	38	20	90.1	89.7	0.4	75.8	77.0	−1.2	80.0	78.0	2.0
11	35	21	55.8	55.7	0.1	70.9	73.0	−2.1	71.2	75.8	−4.6
12	43	21	136.1	142.0	−5.9	79.8	82.7	−2.9	78.3	81.1	−2.8
13	41	21	116.1	110.0	6.1	75.0	73.3	1.7	75.3	72.4	2.9
14	38	21	120.3	122.1	−1.8	87.5	86.9	0.6	88.3	86.8	1.5
15	36	21	97.4	98.4	−1.0	75.2	75.9	−0.7			
16	39	21	95.5	95.6	−0.1	76.4	82.9	−6.5	77.4	83.0	−5.6
17	41	22	131.2	119.9	11.3	76.8	77.8	−1.0	78.4	74.6	3.8
18	40	22	101.0	89.6	11.4	80.4	80.0	0.4	76.7	80.2	−3.5
19	32	24	127.1	143.7	−16.6	76.7	75.8	0.9	81.5	81.0	0.5
20	37	24	130.3	127.1	3.2	79.7	78.6	1.1	80.7	82.9	−2.2
21	38	26	105.7	99.3	6.4	78.3	82.3	−4.0	80.3	83.8	−3.5
22	38	26	130.4	135.2	−4.8	70.5	76.5	−6.0	52.2	69.0	−16.8
Total......	833	449	2257.8	2236.0	21.8	1689.4	1719.0	29.6	1453.9	1497.4	−43.5
Average....	37.9	20.4	102.6	101.6	1.0	76.8	78.1	−1.3	76.5	78.8	−2.3

pairs of classes is so small that we probably should not consider the result representative. Certainly, little if any significance can be attached to the results for Latin, history, and French. The average for the 67 pairs of classes is the weighted average, so that one pair of classes does not influence this average any more than any other pair.

Table VIII presents a summary for those pairs of classes in which the large class is at least twice the size of the small class. By doing this, we are able to examine the achievements of pupils in pairs of classes where the difference in size is marked. The number of classes in any one school subject is so small, with the exception of English, that we are probably not justified in drawing any conclusions for the separate subjects.

Differences in the achievements of paired pupils. Table IX is similar to Table VII, the only difference being that it is based upon the records of only those pupils in the large classes who could be paired with pupils in the corresponding small classes, having the same scores on the general intelligence test. Since the pupils were paired on the basis of their intelligence scores, the average general intelligence of those taken from the large classes would be identical with the average general intelligence of those taken from the corresponding small classes. Hence, the average general intelligence scores are omitted. Table X is similar to Table VIII.

Interpretation of the tables. When all 67 pairs of classes are considered, the average of the general intelligence scores for the large classes is almost identical with that for the small classes, the difference being only one-tenth of a unit. This unit corresponds approximately to one month of mental age. We may, therefore, consider the pupils in the large classes equal in general intelligence to the pupils in the small classes. Both the average "term grade" and the average "examination grade" are slightly larger for the small classes.

The question of the significance of the difference of two averages is involved here. Both examination grades and final grades are known to be highly subjective and to involve a large error which is a combination of a constant error and a variable error. Variable errors tend to offset each other in an average because some of them are negative and some positive. On the other hand, constant errors

TABLE VII. COMPARISON OF AVERAGE INTELLIGENCE SCORES, AVERAGE TERM GRADES, AND AVERAGE EXAMINATION GRADES OF THE DIFFERENT SUBJECTS FOR 67 LARGE AND 67 SMALL CLASSES.

Subject	No. of pairs of classes	Average size		Average inteligence score		Difference	Average term grade		Difference	Average examination grade		Difference
		Large class	Small class	Large class	Small class		Large class	Small class		Large class	Small class	
English	22	37.9	20.4	102.6	101.6	1.0	76.8	78.1	−1.3	76.5	78.8	−2.3
Algebra	18	35.7	20.3	99.6	102.2	−2.6	74.3	76.5	−2.2	76.3	77.2	−0.9
General Science	6	35.7	21.5	110.5	109.6	0.9	75.3	77.2	−1.9	74.8	76.1	−1.3
Botany	4	28.3	18.0	126.3	122.2	4.1	73.8	74.2	−0.4	71.9	72.9	−1.0
Stenography	3	36.7	22.0	83.6	85.9	−2.3	77.6	78.1	−0.5	78.4	80.5	−2.1
Geometry	3	40.7	22.7	128.1	121.1	7.0	75.8	76.6	−0.8	62.9	67.8	−4.9
Arithmetic	3	38.7	22.7	90.2	89.8	0.4	74.7	74.5	0.2	63.3	64.3	−1.0
Typewriting	3	39.3	24.7	83.4	87.9	−4.5	73.2	76.6	−3.4	80.1	83.0	−2.9
French	2	35.0	19.0	110.0	113.2	−3.2	73.7	70.9	2.8	71.2	71.4	−0.2
History	2	33.5	21.0	139.1	137.4	1.7	75.7	74.6	1.1	74.6	74.3	0.3
Latin	1	39.0	25.0	103.7	92.5	11.2	77.3	73.9	3.4	75.4	70.9	4.5
Average		36.5	20.8	104.1	104.0	0.1	75.4	76.7	−1.3	74.8	76.4	−1.6
Total	67											

TABLE VIII. COMPARISON OF AVERAGE INTELLIGENCE SCORES, AVERAGE TERM GRADES, AND AVERAGE EXAMINATION GRADES OF LARGE AND SMALL CLASSES WHERE THE LARGE CLASSES ARE AT LEAST TWICE THE SIZE OF THE SMALL CLASSES.

Subject	No. of pairs of classes	Average size		Average intelligence score		Difference	Average term grade		Difference	Average examination grade		Difference
		Large class	Small class	Large class	Small class		Large class	Small class		Large class	Small class	
English	8	39.1	17.8	90.4	90.6	−0.2	77.1	77.7	−0.6	77.2	78.8	−1.6
Algebra	5	39.6	18.2	96.2	100.4	−4.2	74.6	78.5	−3.9	85.6	86.3	−0.7
General Science	1	42.0	17.0	73.4	78.7	−5.3	70.4	70.0	0.4	70.4	70.0	0.4
Botany	1	33.0	15.0	128.8	117.6	11.2	78.1	71.8	6.3	78.4	75.5	2.9
Stenography	1	36.0	18.0	52.4	55.7	−3.3	73.3	72.9	0.4	71.3	76.5	−5.2
Average		38.9	17.7	91.2	92.4	−1.2	75.7	76.8	−1.1	79.1	80.2	−1.1
Total	16											

31

TABLE IX. COMPARISON OF AVERAGE TERM GRADES AND AVERAGE EXAMINATION GRADES OF PUPILS PAIRED ACCORDING TO EQUAL INTELLIGENCE SCORES IN 67 LARGE AND 67 SMALL CLASSES.

Subject	Number of pairs of classes	Number of pairs of pupils	Average size		Average term grade of paired pupils		Difference	Average examination grade of paired pupils		Difference
			Large class	Small class	Large class	Small class		Large class	Small class	
English	22	336	37.9	20.4	76.4	78.1	−1.7	76.6	78.8	−2.2
Algebra	18	298	35.7	20.3	74.3	76.3	−2.0	76.4	77.1	−0.7
General Science	6	102	35.7	21.5	76.2	78.7	−2.5	75.5	76.6	−1.1
Botany	4	36	28.3	18.0	73.5	72.9	0.6	68.8	70.5	−1.7
Arithmetic	3	56	38.7	22.7	73.8	74.3	−0.5	62.3	64.7	−2.4
Geometry	3	50	40.7	22.7	76.1	77.0	−0.9	61.9	68.2	−6.3
Stenography	3	44	36.7	22.0	76.3	79.8	−3.5	76.5	81.6	−5.1
Typewriting	3	42	39.3	24.7	73.7	76.2	−2.5	77.9	83.6	−5.7
French	2	25	35.0	19.0	71.2	70.0	1.2	68.9	70.0	−1.1
History	2	31	33.5	21.0	78.1	75.4	2.7	75.1	75.4	−0.3
Latin	1	14	39.0	25.0	73.6	79.3	−5.7	73.6	75.0	−1.4
Average			36.5	20.8	75.2	76.8	−1.6	74.4	76.5	−2.1
Total	67	1034								

TABLE X. COMPARISON OF AVERAGE TERM GRADES AND AVERAGE EXAMINATION GRADES OF PUPILS PAIRED ACCORDING TO EQUAL INTELLIGENCE SCORES IN LARGE AND SMALL CLASSES IN WHICH THE LARGE CLASSES ARE TWICE THE SIZE OF THE SMALL CLASSES.

Subject	Number of pairs of classes	Number of pairs of pupils	Average size		Average term grade of paired pupils		Difference	Average examination grade of paired pupils		Difference
			Large class	Small class	Large class	Small class		Large class	Small class	
English	8	116	39.1	17.8	76.6	77.7	−1.1	76.7	78.5	−1.8
Algebra	5	84	39.6	18.2	73.7	77.5	−3.8	83.6	86.3	−2.7
General Science	1	16	42.0	17.0	71.9	71.9	0.0	71.9	71.9	0.0
Stenography	1	11	36.0	18.0	72.7	72.7	0.0	70.6	77.6	−6.0
Botany	1	11	33.0	15.0	74.6	71.8	2.8	71.8	75.5	−3.7
Average			38.9	17.7	75.0	76.6	−1.6	77.9	80.3	−2.4
Total	16	238								

are not eliminated in averages. The number of pupils included in the 67 pairs of classes is sufficiently large so that the average grades include a variable error which is probably so small as to be negligible. For example, if we assume that the probable variable error of a grade is as much as 10 points, which is probably in excess of the actual probable variable error, the probable variable error of the average for the small classes would be less than three-tenths of one point. In the case of the large class, it would be materially less than in the small class. The constant errors, which are expressed in the tendency of some teachers to give higher grades than others, are probably included in both groups in approximately the same proportion, since the same teacher assigned grades to both a small class and a large class. If this is true, a difference of 1.3 between the average term grades is probably significant, although there is a reasonable doubt. This doubt is materially increased and probably the difference loses its significance when we recall that the achievements of the two groups of students were not measured at the beginning of the experimental period.

When we consider the summaries for the different subjects, we find the differences between the average achievements of the pupils in the two types of classes materially larger in a number of cases than the differences between the averages for the 67 pairs of classes. However, in interpreting these differences it is necessary to remember that the number of pupils is materially less and, hence, a difference must be materially greater in order to be significant. It is perhaps significant that negative differences predominate. This suggests that the average achievements in the small classes are somewhat superior to those in the large classes, but in drawing conclusions from this condition it is necessary to bear in mind the fact that the achievements of the students were not measured at the beginning of the experimental period.

When we turn to Table VIII, which is restricted to those classes in which the large class is at least twice the size of the small class, we find that the pupils in the small classes were slightly superior in general intelligence. If the size of class is a potent factor in determining the achievements of pupils, we should naturally expect to find a greater difference in the achievements of the two types of classes than we found in Table VIII. The fact that the difference is less, both absolutely and relatively, suggests that the size of class is not a potent factor in determining the achievements of pupils. However, the number of pairs on which Table VIII is based is so small that no great importance should be attached to this observation.

When we examine Table IX, we find that the differences between the averages for the 67 pairs of classes are only slightly larger than those given in Table VII. In general, there are few significant differences to be noted in a comparison of these two tables. One of the most significant is the reversal of the relative achievements of the large class and the small class in Latin. In Table VII, when all pupils were considered, those taught in the large class were shown to be distinctly superior in achievement. When only the paired pupils were considered, those taught in the small class were distinctly superior in achievement. Considering the table as a whole, we are justified in asserting that it tends to corroborate the interpretations suggested for Table VII.[4]

Conclusion: relation of size of class in high school to school efficiency. The tables of this chapter show that at the end of the experimental period the achievements of the students in the two types of classes were approximately equal, and there is a slight indication that those taught in small classes were superior. Since the educational investment can be materially decreased by increasing the size of class in the high school, one might infer that the efficiency of the school would be increased by organizing classes enrolling from 35 to 40 students instead of classes enrolling from 20 to 25. In addition to the fact that there are several uncontrolled factors whose influence is unknown, it is necessary to bear in mind the exact conditions of the experiment. Since the same teachers taught both a small class and a large class, there was no difference between the total amount of work done by the teachers who handled the large classes and the teachers who handled the small classes. In fact, they were the same teachers. Thus, this experiment failed to set up the conditions of large classes as a general plan of organization of a high school. It did, however, realize the conditions which not infrequently exist in the smaller high schools where it is desirable to have a few large classes assigned to teachers who are given compensating small classes or who have the number of classes reduced accordingly. The results of the experiment, therefore, can be applied only to those situations in which the teaching load is kept constant. In such cases the evidence collected indicates that approximately the same average achievement can be expected from the pupils taught in large classes as from those taught in small classes. In other words, the results of this experiment indicate that there is no loss of efficiency caused by organizing a few large classes if the other work assigned to the teacher is such that the teaching load is not increased.

---

[4] In his report, Mr. Stevenson attempted the further analysis of the data by ascertaining the percent of pairs of pupils in which the pupil in the small class received a higher grade than the corresponding pupil in the large class. When all pairs of pupils are considered, he shows that only in 40 percent of the pairs does the pupil in the small class surpass his mate in the large class. However, it is impossible to draw any conclusion from this fact, because we do not know the percent of pairs in which the two pupils received the same mark. Since, in a majority of the cases, the grades were reported in terms of letters and only five marks were recognized, it is reasonable to expect that in a relatively large percent of the cases both of the paired pupils received the same final grade.

One should recognize that the results of this experiment should not be applied to the question of the size of class where increasing the size of class results in a distinct increase in the teaching load. The instruction which students receive is given partly in the class room and partly through written work and individual conferences. In such subjects as English composition, algebra, and science requiring laboratory work, it is customary with most teachers to require a large amount of written work. A teacher who gives instruction to five classes of 40 students each has a much heavier teaching load than the teacher who instructs five classes of 20 students each, unless he introduces compensating changes in the amount of written work, in the method of handling it, and in the number of individual conferences. In such cases the question of class size is so intimately connected with the method of instruction that we are not justified in drawing any inferences from an investigation in which the method of instruction was assumed to be the same for both types of classes.

CHAPTER V

SUGGESTIONS FOR EDUCATIONAL EXPERIMENTATION[1]

The two studies described in Chapters III and IV make very slight contributions to a scientific determination of the relation between the size of class and the efficiency of a school system. They do, however, seem to the writer of this chapter suggestive with reference to the procedure of educational experimentation. The causes of the failure of these studies to produce reliable and significant results have been mentioned in the two preceding chapters but they may be summarized under two general heads: *first,* failure to set up and maintain appropriate experimental conditions and, *second,* the lack of adequate instruments for measuring the achievements of pupils.

A prerequisite for setting up and maintaining appropriate experimental conditions is a complete analysis of the problem being studied. The various factors involved must be recognized by the experimenter and the possibility of any relations which may exist between these factors must also be considered. For example, in the high school many factors contribute to the achievements of students, or the educational output of the school. In addition to the size of class, which is the factor whose relationship to school efficiency was studied, it is necessary to recognize methods of instruction, the personality and enthusiasm of the teacher, the discipline of the class and of the school, the general spirit of the school, the general attitude of the community toward the school, the time of day when the class recites, the textbooks used, the equipment, including the building, the "spiritual composition" of the class, the general intelligence of the students, their nationality, their past experience, both in school and out of school, the stage of advancement reached in their education, and possibly even other factors. It appears likely that certain of these factors are interrelated. The size of class is likely to affect the enthusiasm of the teacher, particularly if it determines the instructional load carried by the teacher. It also appears that some relationship exists between the size of class and methods of instruction, and between the size of class and discipline. The existence of a functional relationship between two or more factors makes it impossible under normal conditions to produce variations in one factor without, at the same time, causing changes in the others. Failure to analyze the problem sufficiently will frequently cause the results of an educational experimentation to have little significance, and consequently the time and money invested in the study will be largely wasted.

---

[1] This chapter has no counterpart in the report submitted by Mr. Stevenson. It is entirely the contribution of the present director of the Bureau of Educational Research.

When one considers the total product of the educational process one cannot fail to become impressed with the inadequacy of our present educational tests as instruments for the measurement of the various elements of this product. In the study relating to the size of class in the elementary school a more elaborate group of tests might have been used, but even if this had been done, it does not appear likely that one would be justified in asserting that the total product of education had been measured. In the high school no standardized educational tests were used. It was attempted to have the achievements measured by means of a final examination and by the term "grade" given to the students. The writer of this chapter is not aware of the considerations which resulted in the decision not to use any of the standardized educational tests that have been devised in the field of high school subjects, but it is likely that this decision was due to the belief that none of the available educational tests were sufficiently satisfactory measuring instruments to justify their use in this investigation. The present writer is inclined to share this belief. Thus, we cannot escape the conclusion that at the present time we do not have available instruments for measuring the outcomes of teaching which permit reliable educational experimentation when it is necessary to measure the total product of instruction.

Incidentally, attention may be called to the fact that more consideration should be given to the errors involved in the data and to the effect of these errors upon the results of statistical calculations. For example, it is highly important to know what significance should be attached to a difference between two averages.

In view of the difficulties encountered in setting up and maintaining appropriate experimental conditions and in view of the imperfections and limitations of our present educational tests, it is not inappropriate to question the wisdom of undertaking such complex educational experimentation as has been described in this monograph. It is true that there are many educational problems which are highly important. For example, an increase in the size of class would result in a material reduction in the educational expenditures for instruction. A reliable scientific determination of the relationship existing between the size of class and school efficiency would be a valuable contribution, but it is doubtful whether such a determination of this relation is at the present time possible.

To the present writer, it appears highly important that those engaged in educational research should give very careful consideration to the sort of problems to which they devote their energies. It is, of course, necessary that pioneer work be done, and in studies of this type, it is not always possible to anticipate the limitations of one's procedure. As a result it may become necessary to "scrap" a project because of the defects that appear in the course of one's work. Such losses are unavoidable in extending the frontiers of educational research. When an investigation is not pioneer work an experimenter should determine the limitations of his procedure in advance, and unless it appears likely that information of considerable value will be secured in spite of the limitations the investigation should not be undertaken. The fact that a problem is important does not justify its study. Educational experimentation which involves the use of faulty method and faulty instruments not only fails to make adequate contributions to our educational progress, but, more important, it tends to reflect unfavorably upon the application of the methods of research to the field of education.

BULLETIN NO. 11 1.

BUREAU OF EDUCATIONAL RESEARCH
COLLEGE OF EDUCATION

RELATION OF SECTIONING A CLASS TO THE EFFECTIVENESS OF INSTRUCTION

by

WALTER S. MONROE.

PRICE 15 CENTS

PUBLISHED BY THE UNIVERSITY OF ILLINOIS, URBANA
1922

UNIVERSITY
of ILLINOIS
PRESS

PREFACE

The educational experiment reported in this bulletin was initiated by the former Director of the Bureau of Educational Research and the data collected under his supervision. The present Director of the Bureau is responsible for the tabulation of the data and for the preparation of this report.

This investigation was made possible through the cooperation of Superintendent Peter A. Mortenson and of certain principals and teachers of the Chicago Public Schools. Not only did they cooperate in the collection of the data but they also made substantial contributions to the project by supplying test materials. The writer is glad to acknowledge the indebtedness of the Bureau of Educational Research to all who contributed to this project.

WALTER S. MONROE, *Director*

November 10, 1922

Relation of Sectioning a Class to the Effectiveness of Instruction

The problem. The purpose of this educational experiment was to determine the relative effect upon the achievements in certain school subjects of three plans of sectioning a class. A "class" is defined as the total number of children assigned to a teacher for instruction even though they may be divided into two or more groups for instructional purposes. The three plans of sectioning a class considered in this investigation are: (1) teaching a class as a single unit; (2) dividing the class into two equal groups approximately equivalent with respect to general intelligence; (3) dividing the class into three equal groups approximately equivalent with respect to general intelligence. When a class is taught as one group, all of the pupils recite at the same time. Following the recitation there is a period for study. Thus under this plan the work of the teacher alternates between "hearing classes" and supervising the study of the pupils. When a class is taught as two sections, one group recites while the other group studies. In this case the teacher's time is almost wholly devoted to "hearing classes." Any supervision of the study of the pupils is of necessity given incidentally and at irregular intervals when the teacher is fortunate enough to have a few minutes of leisure during a recitation period. When a class is divided into three sections, the conditions are much the same except that necessarily the length of the recitation periods is reduced. In general pupils of one section study during the recitation periods of the other two sections.

The specific problem of this investigation was to determine the relative effect of these three plans of sectioning a class upon the direct results of instruction in certain school subjects. In other words this investigation sought to answer the question, "Which is the best plan of sectioning a class?"

General plan of the experiment. If it were possible to secure three groups of classes so that all factors which affect the results of instruction were equivalent in the beginning of the experiment and could be controlled throughout the experimental period, the simplest procedure would be to have one group of classes taught as a unit, another group taught in two sections and a third group in three sections. However, it would be difficult, if not impossible, to secure exact equivalence of teaching ability and of pupil material. Our facilities for measuring the ability of teachers are extremely crude and at best it would be difficult to demonstrate that any differences found in the results of instruction were not produced largely by differences in teaching ability. It is true that we have a number of general intelligence tests which might be used to measure the quality of the pupil material. However, the limitations of these instruments are such that one would be unable to interpret small differences in the resulting achievements.

In order to avoid these two difficulties this experiment was planned so that the same teacher should instruct a given class when organized according to two different plans of sectioning. This, necessarily, must be done during successive semesters. This procedure insured the constancy of the teacher, although not necessarily of teaching ability since the ability of a given teacher may vary from semester to semester with different types of class organization. In order that the pupil material might be the same for the two plans of class organization one hundred percent promotion was secured at the middle of the school year. Thus, a teacher who instructed a class as one section during the first semester of this experiment instructed the same pupils during the second semester but with the class divided into two or three sections. Other teachers taught classes organized according to other combinations of sectioning.

This general plan of the experiment makes the semester a variable factor. It is possible that pupils may normally make greater progress during one semester than during the other. Furthermore, the gain of second trial scores over first trial scores is likely to be much greater than the gain of third trial scores over second trial scores simply because the pupils become acquainted with the testing procedure. In order to balance these two variable factors it was necessary to arrange experimental groups in pairs. Thus, corresponding to an experimental group of classes which was taught as a single section during the first semester and as three sections during the second semester, there was another group of classes taught as three sections during the first semester and as a single section during the second semester. In dividing a class into sections the scores yielded by the general intelligence tests were used to secure sections of approximately equivalent pupil material. Six experimental groups of classes were organized as follows:

Group I. Classes taught as a single section during the first semester and as three sections during the second semester.

Group II. Classes taught as three sections during the first semester and as one section during the second semester.

Group III. Classes taught as one section during the first semester and as two sections during the second semester.

Group IV. Classes taught as two sections during the first semester and as one section during the second semester.

Group V. Classes taught as two sections during the first semester and as three sections during the second semester.

Group VI. Classes taught as three sections during the first semester and as two sections during the second semester.

So far as the writer knows, essentially the same methods of instruction and subject-matter were followed in all of these groups. The investigation was confined to Grades II, V, and VII in order to reduce the labor and expense. As these grades are fairly representative of the three divisions of the elementary school, primary, intermediate and grammar, it is not likely that different results would be obtained in the other grades. The number of classes, the total enrollment, and the number of complete records in each experimental group are given in Table I.

TABLE I. NUMBER OF CLASSES, TOTAL ENROLLMENT, AND NUMBER OF COMPLETE RECORDS IN EACH OF THE EXPERIMENTAL GROUPS

Grade		Group						Total
		I	II	III	IV	V	VI	
II	Number of classes	7	4	3	6	7	3	30
	Total enrollment	348	201	138	288	324	162	1461
	Complete records	240	111	103	208	224	89	975
V	Number of classes	2	2	8	4	4	4	24
	Total enrollment	87	92	379	192	196	181	1127
	Complete records	70	72	326	133	157	143	901
VII	Number of classes	3	3	5	5	2		18
	Total enrollment	141	140	244	214	91		830
	Complete records	119	109	186	159	86		659

The data collected. Through the cooperation of Superintendent Peter A. Mortenson of the Chicago Public Schools and of certain principals and teachers, the Bureau of Educational Research carried on this investigation during the school year of 1920-21. Experimental classes were organized in sixteen elementary schools.[1] For measuring the general intelligence of the pupils the Pressey Primer Scale was used in the second grade, and the Illinois General Intelligence Scale in the other two grades. The achievements of the pupils in the second grade were measured by means of the Pressey Scale of Attainment No. 1. In the fifth and seventh grades achievements were measured by Monroe's Standardized Silent Reading Tests, Revised, Monroe's General Survey Scale in Arithmetic, and Buckingham's Problem Scale in Arithmetic, Divisions 1 and 2. The general intelligence tests were given only at the beginning of the experiment, October 11, 1920. Form 1 of the achievement tests was given at this time. Form 2 of the achievement tests was administered at the close of the first semester, February 3, 1921. At the close of the experimental period, May 11, 1921, Form 1 was again given.

The tests were administered by the teachers who also scored the test papers and entered the scores upon individual record cards. This, however, was done only after all of the teachers involved in the experiment had been called together for the purpose of acquainting them with the tests. In this explanation several tests were administered to the teachers in exactly the same way as they were to be administered to the pupils. In addition detailed instructions were supplied to the teachers for all steps of the work. Since no comparisons were made between the scores yielded by tests administered by different teachers it is felt that this procedure in the administration of the tests does not seriously affect the results of the experiment.

Limitations of the experiment to be kept in mind in interpreting the results. A number of conditions must be kept in mind in interpreting the results. In the first place practically all of the teachers who cooperated in the investigation had been accustomed to teaching classes in two sections. A few, perhaps 1 in 20, had taught a class as a single section but, so far as the writer was informed, no teacher had had any experience in instructing a class in three sections. Thus, it is altogether likely that most of the teachers had acquired a technique of instruction which would prove more successful with a class divided into two sections than with a class divided into either one or three sections. Furthermore, there appears to be a prejudice against the division of a class into three sections. Thus, there is introduced a factor which may be expected to produce greater achievements in classes taught as two sections than in classes taught as either one or three sections. The effect of this factor is, however, unknown but it should by all means be recognized in interpreting the results.

---

[1] These sixteen schools were the following: Brown, Dante, Douglas, Fiske, Jenner, Julia Ward Howe, Morse, Otis, Pullman, Scanlan, Shields, Spry, Van Vlissingen, Ward, Wentworth, and West Pullman.

The instruments used for measuring the achievements of the pupils do not measure all achievements resulting from instruction. They can be considered to do no more than measure representative samples of the achievements within their respective fields. Outside of silent reading and arithmetic, in which tests were given, there are many important achievements of which no attempt was made to secure direct measurements. It is, of course, possible that the measures of achievements secured correlate closely enough with all other achievements resulting from instruction, that a sufficiently accurate index of all achievements is furnished for judging the relative effectiveness of the instruction in the different experimental groups. However, convincing experimental evidence on the point is wanting and, for this reason, due caution must be exercised in extending the conclusions of this experiment to school subjects other than silent reading and arithmetic, as well as to the more subtle outcomes engendered by the social contacts of the school room.

Finally, it must be remembered that this investigation was carried on in classes enrolling approximately 45 pupils. Hence it does not necessarily follow that the conclusions would apply to classes enrolling 20 to 30 pupils. It is possible that this change in the size of class might produce a complete reversal in the conclusions.

Method of summarizing data. After rejecting records which were incomplete and obviously inaccurate, the scores yielded by an application of a test were combined in a total distribution for each experimental group. Thus, a distribution was formed of the first trial scores made on Monroe's Standardized Silent Reading Tests, Revised, by the group of fifth grade pupils enrolled in "classes taught as a single section during the first semester and as three sections during the second semester." In the same way distributions of scores were formed for each of the experimental groups and for each application of the test. The gain in achievement during the first semester was found by subtracting the average score for the first trial of a test from the average score of the second trial. The gain for the second semester was found by subtracting the average score of the second trial from that of the third trial. A second measure of gain was secured by following a similar procedure with the median scores but these gains are not given in this report as they were, in general, in agreement with those calculated from the average scores.

In calculating these gains no account was taken of the possible non-equivalence of the different forms of the tests used. In fact no accurate information concerning the equivalence of duplicate forms is available except for Monroe's Standardized Silent Reading Tests, Revised, and for Monroe's General Survey Scale in Arithmetic. The duplicate forms of these two tests have been shown to be approximately equivalent.[2] However, since Form 1 of each test was used twice and the average scores calculated from it were used both as subtrahends and minuends, and since the gain for any plan of sectioning is computed from both semesters the non-equivalence of Forms 1 and 2 of the tests used will not affect the comparisons of gains made in the following tables.

The point scores yielded by the different tests are expressed in terms of different units and from different zero points. Thus before any combination from the results of the different tests can be made it is necessary to express the gains in terms of a common unit. The usual assumption in such cases is that the standard deviation of the distribution of scores represents the same increment of ability for one test as for another. On the basis of this assumption a total distribution for each test was secured by adding the distributions of the six experimental groups within a grade. This was done for the scores secured at each period of testing. The average of the three standard deviations was assumed to represent the same increment of ability for each test and was used as a divisor for reducing the gains to the basis of a common unit. For example, during the first semester the fifth grade pupils in Group I classes made a gain in arithmetic of 23.82 points. During the second semester they made a gain of 21.5 points. The average standard deviation of the arithmetic scores in the fifth grade is 19.65. Using this as a divisor we secure as quotients 1.21 and 1.09. In this manner the entries in Tables II, III and IV were obtained. The two quotients whose calculation was explained are given in Table III.

Tables II, III and IV are similar in structure and are to be read in the same way. The gains for the different experimental groups

---

[2] Monroe, W. S. Illinois Examination, University of Illinois Bulletin Vol. 19, No. 9, Bureau of Educational Research Bulletin No. 6. Urbana: University of Illinois, 1921. 70 p.

TABLE II. GAINS IN ACHIEVEMENT MADE IN THE DIFFERENT EXPERIMENTAL GROUPS IN THE SECOND GRADE*
(GAINS COMPUTED FROM AVERAGE SCORES)

Group	No. of Pupils	Test 1			Test 2			Test 3			Test 4			Average		
		Gain	Gain	Dif.	Gain	Gain	Dif.	Gain	Gain	Dif.	Gain	Gain	Dif.	Gain	Gain	Dif.
		(1)	(3)		(1)	(3)		(1)	(3)		(1)	(3)		(1)	(3)	
I	240	1.42	.55		.94	.45		1.49	.27		1.37	.71		1.31	.50	
II	111	.90	1.11		.62	.82		.69	1.17		.58	1.18		.70	1.07	
Average		1.16	.83	.33	.78	.64	.15	1.09	.72	.37	.98	.95	.03	1.00	.78	.22
		(1)	(2)		(1)	(2)		(1)	(2)		(1)	(2)		(1)	(2)	
III	103	1.10	.52		.63	.60		1.50	.51		1.39	.47		1.16	.53	
IV	208	.87	1.35		.56	1.03		.48	1.20		.55	1.21		.62	1.20	
Average		.99	.94	.05	.60	.82	−.22	.99	.86	.14	.97	.84	.13	.89	.86	.02
		(2)	(3)		(2)	(3)		(2)	(3)		(2)	(3)		(2)	(3)	
V	224	1.41	.56		.63	.56		1.34	.27		1.53	.62		1.23	.50	
VI	89	.04	1.56		1.01	.42		.28	1.36		.31	1.25		.41	1.15	
Average		.73	1.06	−.34	.82	.49	.33	.81	.82	−.005	.92	.94	−.02	.82	.83	.006

*The numbers in parentheses in the body of the table indicate the number of sections in which the classes were taught.

TABLE III. GAINS IN ACHIEVEMENT MADE IN THE DIFFERENT EXPERIMENTAL GROUPS IN THE FIFTH GRADE*
(GAINS COMPUTED FROM AVERAGE SCORES)

Group	No. of Pupils	Reading Rate			Reading Comprehension			Arithmetic			Arith. Problems 1st Division			Arith. Problems 2nd Division			Average		
		Gain	Gain	Dif.	Gain	Gain	Dif.	Gain	Gain	Dif.	Gain	Gain	Dif.	Gain	Gain	Dif.	Gain	Gain	Dif.
		(1)	(3)		(1)	(3)		(1)	(3)		(1)	(3)		(1)	(3)		(1)	(3)	
I	70	.68	-.39		.30	.11		1.21	1.09		.35	.13		.65	.35		.80	.32	
II	72	.55	.09		.58	.37		.28	.75		.24	.40		-.14	.95		.38	.64	
Average		.62	-.15	.77	.44	.24	.20	.75	.92	-.18	.30	.27	.03	.26	.65	-.40	.59	.38	.11
		(1)	(2)		(1)	(2)		(1)	(2)		(1)	(2)		(1)	(2)		(1)	(2)	
III	326	-.12	.83		.29	.63		.61	.32		.79	.18		.84	.11		.60	.52	
IV	133	.57	-.16		.30	.25		.51	.87		.68	-.05		.25	.24		.58	.29	
Average		.23	.34	-.11	.30	.44	-.15	.56	.60	-.04	.74	.07	.67	.55	.18	.37	.59	.40	.19
		(2)	(3)		(2)	(3)		(2)	(3)		(2)	(3)		(2)	(3)		(2)	(3)	
V	157	.23	.69		.45	.66		.54	.29		.45	.20		.79	.16		.62	.50	
VI	143	1.03	-.23		.73	.03		.27	.24		.14	.70		.04	.68		.55	.36	
Average		.63	.23	.40	.59	.35	.25	.42	.27	.14	.30	.45	-.16	.42	.42	-.005	.58	.43	.16

*The numbers in parentheses in the body of the table indicate the number of sections in which the classes were taught.

TABLE IV. GAINS IN ACHIEVEMENT MADE IN THE DIFFERENT EXPERIMENTAL GROUPS IN THE SEVENTH GRADE*
(GAINS COMPUTED FROM AVERAGE SCORES)

Group	No. of Pupils	Reading Rate			Reading Comprehension			Arithmetic			Arith. Problems 1st Division			Arith. Problems 2nd Division			Average		
		Gain	Gain	Dif.	Gain	Gain	Dif.	Gain	Gain	Dif.	Gain	Gain	Dif.	Gain	Gain	Dif.	Gain	Gain	Dif.
		(1)	(3)		(1)	(3)		(1)	(3)		(1)	(3)		(1)	(3)		(1)	(3)	
I	119	−.60	.13		.47	.93		.14	.23		.34	−.03		.37	.17		.18	.36	
II	109	.31	−.41		.07	.22		.71	.18		−.20	.58		.07	.78		.24	.34	
Average		−.15	−.14	−.005	.27	.58	−.31	.43	.21	.22	.07	.28	−.21	.22	.48	−.26	.21	.39	−.14
		(1)	(2)		(1)	(2)		(1)	(2)		(1)	(2)		(1)	(2)		(1)	(2)	
III	186	−.14	.60		.32	.41		.31	.92		.48	.04		.33	.03		.33	.50	
IV	159	.45	.28		.41	.91		.51	.26		−.07	.53		.23	.08		.38	.52	
Average		.16	.44	−.29	.37	.66	−.30	.41	.59	−.18	.21	.29	−.08	.28	.06	.23	.35	.51	−.15

*The numbers in parentheses in the body of the table indicate the number of sections in which the classes were taught.

are arranged in pairs. In Table II, the gain for Group I on Test 1 when taught in classes of one section is 1.42. When taught in three sections the gain is .55. The gain for Group II classes when taught in one section is .90 and when taught in three sections it is 1.11. The Group I classes were taught in one section during the first semester but the Group II classes were taught in one section during the second semester. This difference in time is largely responsible for the differences in the size of the gains.

Interpretation of results. In interpreting the gains in Tables II, III and IV it is necessary to keep in mind both the constant and variable errors of measurement which are involved in the original data as well as the chance variations in the gains due to sampling. The variable errors of measurement in the original data depend upon the reliability of the tests used. If we assume a coefficient of reliability[3] of .84 for Test 1, it can be shown that the probable variable error of measurement is approximately .25 when expressed in terms of sigma which is the unit used in expressing the gains in Tables II, III, and IV.[4] A probable error of measurement of .25 means that the scores for 50 percent of the pupils involve variable errors which are less than .25. For the other 50 percent the variable errors will be greater than .25. The presence of variable errors of measurement affects the average of the scores as shown by the following formula in which N is the number of scores upon which the average is based.

$$\text{P. E.}_{\text{M average}} = \frac{\text{P. E.}_{\text{M}}}{\sqrt{N}}$$

Substituting in this formula for Group I, we find the probable error of measurement of the average ($\text{P. E.}_{\text{M average}}$) is .017; for Group II it is .024. The gain 1.42 is the difference between the two averages.

---

[3] The coefficient of reliability assumed here is probably higher than would be found for this test. When based upon the scores of a single grade, the coefficient of reliability for Monroe's General Survey Scale in Arithmetic is approximately .85. For Monroe's Standardized Silent Reading Test 1, Revised, the coefficients of reliability are approximately .75 for rate and .65 for comprehension. For Test II they are about .08 higher. The reliability of the other tests is not known.

[4] The formula for the probable variable error of measurement is

$$\text{P. E.}_{\text{M}} = .6745\,\sigma\sqrt{1-r_{12}}$$

In this case $\sigma = 1$.

The probable error of the difference of the two averages is given by the following formula

$$\text{P. E.}_{\text{Dif.}} = \sqrt{\overline{\text{P. E.}_{1}}^{2} + \overline{\text{P. E.}_{2}}^{2}}$$

In this formula P. E.$_1$ and P. E.$_2$ stand for the probable errors of measurement of the two averages whose difference is taken. In this case P. E.$_1$ is equal to P. E.$_2$ since we have used the average of the standard deviations of the several distributions in reducing the gains to a comparable basis. Applying the above formula, we find that the probable variable error of measurement to be associated with 1.42 is .024 and with .90 is .034. The formula for the probable error of the sum of the two averages is the same as that for their difference. Hence we may calculate the probable error of measurement to be associated with the average gain 1.16 by taking one half of the probable error of measurement of the sum of the two averages. The P. E.$_M$ of the average gain 1.16 is .020.

Since the probable variable error of measurement depends only upon the magnitude of the standard deviation of the scores and the number of scores, we will obtain the same result for the gains of these two groups when taught in classes of three sections. The probable variable error of measurement of the difference (.33) may be calculated by the formula given above. It is .028.

This probable variable error of measurement is relatively small in comparison with the gain .33, and in general when an average or difference is three or four times its probable error it can be considered significant. Hence, if we had to consider only the variable errors of measurement we would be justified in asserting that this difference was significant and could not be due to the presence of these errors in our original data. However, it should be remembered that we have been liberal in the estimate of the coefficient of reliability. It is likely that the true value of the probable error is much larger.

Since all gains are expressed in terms of a common unit the probable variable errors of measurement found for the entries under Test 1 will apply also to Tests 2, 3, and 4 provided we assume the same coefficient of reliability for these tests. The probable variable error of measurement of the average is affected by the number of cases from which the average is computed. Hence for the gains made by other groups it will be slightly greater, since the number of scores is smaller for those groups. In Table III the number of scores in Groups III and IV is slightly larger. Hence a smaller probable variable error of measurement will be found, but for all of the other groups it will be larger than the one which we have considered in detail. In several cases the difference in gains is so small that when compared with the probable variable error of measurement it cannot be considered as significant.

In addition to the variable errors of measurement, it is necessary to consider the chance variations in the gains due to sampling even when the sample has been chosen without bias. The probable error of an average due to sampling is given by the following formula

$$\text{P. E.}_s = .6745 \frac{\sigma_{\text{dist.}}}{\sqrt{N}}$$

Since sigma (σ) has been used as a unit in terms of which the gains are expressed, $\sigma_{\text{dist.}}$ equals 1 for our calculations.[5] In the case of Group I, $\text{P. E.}_s = .044$. The gain 1.42 is the difference between two averages and hence it would be necessary to apply the formula for the probable error of the difference of the two averages. This being done we find that the P. E._s to be applied to the gain (1.42) is .062. In case of Group II, $\text{P. E.}_s = .064$ and for the difference between the two averages it is .090. For the average 1.16, $\text{P. E.}_s = .055$. For the difference .33, $\text{P. E.}_s = .078$.

When we consider the probable error due to sampling (.078) in addition to the probable variable error of measurement (.028) the difference (.33) would probably be significant and indicate a slight superiority in achievement as measured by Test 1 for the pupils taught in classes of one section, provided no other errors could be considered to affect this difference. It is, however, necessary to consider the constant errors of measurement. Their exact magnitude can not be known but their presence is evident. For example, in Table II the gains on Test 1 for Groups I and II when taught as one section are 1.42 and .90 respectively. The gain of 1.42 was made during the first semester and is the difference between the first and second trial scores. The gain of .90 was made during the second semester and is the difference between the second and third trial scores. Due to the pupils becoming acquainted with the tests and

---

[5] This is not the true value of σ. The variable errors of measurement tend to increase the value of the obtained sigma. The relation is given by the formula

$$\sigma_{\text{true}} = \sigma_{\text{obtained}} \sqrt{r_{12}}$$

the testing procedure, both of these gains involve a constant error. This tends to make the obtained gain larger than the true gain, but as the practice effect of the second trial scores over the first trial scores is larger than that of the third trial over the second trial scores, it is reasonably certain that the gain for Group I (1.42) contains the larger constant error. The gains made by these two groups when taught in classes of three sections are .55 and 1.11. Both of these gains involve a constant error but in this case the larger constant error is found in the gain for Group II. Each of the average gains for these two groups (1.16 and .83) includes a relatively large constant error but the two errors are much more nearly equal than those included in the gains for each group separately. Hence, we are probably justified in considering their difference (.33) to be relatively unaffected by the presence of constant errors in any of our original data.

However, the neutralization of the constant errors which seems plausible, if not probable, in the case we have just considered does not appear to have taken place in a number of the other differences in this group of tables. With the exception of Groups I and II in Table II some of the differences are positive but others are negative for each pair of groups, although it is not impossible that a given plan of sectioning a class might be more effective in one subject than in another. The variations in the signs of the differences do not appear to occur in such a way as to justify this explanation of the negative gains. It is likely that a constant error was introduced in certain groups of scores which was not neutralized in the difference. For example, Group VI is shown by Test 2 to have made a larger gain during the second semester when taught in two sections. Each of the other tests shows a smaller gain for this semester and this we should expect as the gain is the difference between the second and third trial scores. The probable explanation of this condition is that in some way a constant error was introduced in one set of scores yielded by Test 2 for Group VI. An examination of Tables III and IV reveals several similar instances. Hence, we are forced to the conclusion that at least certain sets of scores involve an unknown constant error. The fact that this happened in certain cases tends to make one suspicious of the presence of an unknown constant error in other sets of scores even though evidence of its presence is lacking.

It is perhaps significant that in the case of the differences in gains between classes taught as one section and classes taught in three sections, eight gains are positive while six are negative. The same situation prevails with respect to the gains made by classes taught in one section when compared with the gains made by classes taught in two sections. For classes taught in two sections compared with classes taught in three sections, we have records only in the second and fifth grades. Four of the differences are positive while five are negative.

Conclusion. The facts presented in Tables II, III, and IV and the errors they include appear to justify the conclusion that there is no evidence of greater achievements being made by pupils when taught in classes organized on the basis of one plan of sectioning than in classes organized on a different plan of sectioning. Since the teachers were more experienced in teaching classes in two sections and probably preferred this plan of organization this condition might appear to mean that the division of classes into two sections was the least efficient of the three plans. However, in the writer's judgment this conclusion is not justified. The most obvious inference, in his opinion, to be drawn from the data of this experiment is that the educational tests used do not yield sufficiently accurate and precise measures of achievement to make possible the determination, under the conditions of this experiment, of the best method of sectioning a class. It is likely that the differences in the gains made during a period of less than a semester are not large. This being the case it is necessary either to extend the experimental period or to secure more precise measures of achievement. The magnitude of the probable variable error of measurement of the difference and also of the probable error due to sampling can be decreased by increasing the number of pupils in the experimental groups, but the constant errors are not affected by any increase in the number of cases. Certain constant errors are neutralized in the differences but, as we have shown, other constant errors which occur in only certain sets of scores were not eliminated. The presence of these constant errors is due to imperfections in the educational tests used. Therefore, it appears that until our instruments for measuring achievements of school children are materially improved we cannot expect such educational experiments as the one described in this report to lead to reliable conclusions.

BULLETIN NO. 12

BUREAU OF EDUCATIONAL RESEARCH
COLLEGE OF EDUCATION

THE USE OF INTELLIGENCE TESTS AS A BASIS OF SCHOOL ORGANIZATION AND INSTRUCTION

by

CHARLES W. ODELL
Associate, Bureau of Educational Research

PRICE 50 CENTS

PUBLISHED BY THE UNIVERSITY OF ILLINOIS, URBANA

1922

UNIVERSITY
of ILLINOIS
PRESS

TABLE OF CONTENTS

PAGE

Preface 5

I. The Plan and Conduct of the Experiment 7

II. Conditions at the Beginning of the Experiment 20

III. The Efficiency of the Two Groups of Schools as Measured by the Rates of Progress of the Pupils 27

IV. The Efficiency of the Two Groups of Schools as Measured by the Achievements of the Pupils 41

V. A Study of the Pupils Who Remained in School Throughout the Course of the Experiment 56

VI. A Special Study of the Brighter and Duller Pupils 60

VII. Results and Conclusions 64

Appendix A. A Comparison of the Pupils Entering and Leaving School During the Experiment with the Total Number of Pupils 69

Appendix B. The Reliability and Correlation of the Tests Used in this Experiment 72

Appendix C. The Omnibus Test 78

PREFACE

How to use most effectively the information yielded by general intelligence tests is one of the most important questions before the educational world at the present time. Many of our leading educational thinkers are urging that the children in our elementary schools be grouped into grades on the basis of their mental ages and divided into sections within the grade on the basis of intelligence quotients. Other educators maintain that this should not be done. In this monograph Dr. C. W. Odell presents the results of an investigation extending over nearly two years in which he has studied with unusual care certain of the questions involved in the proposal that we reorganize our schools on the basis of the results yielded by general intelligence tests. Because the questions studied are highly important it is felt that a somewhat detailed report is justified. In order to assist the reader in understanding the experiment the organization of the experimental schools has been described in detail.

This investigation was undertaken at the invitation of Superintendent Peter A. Mortenson of Chicago. Its execution was made possible by the cooperation of Assistant Superintendent A. B. Wight and of certain principals and teachers in the Chicago public schools. To all who have cooperated in the course of the investigation the Bureau of Educational Research desires to acknowledge its indebtedness.

Walter S. Monroe, *Director.*

November 10, 1922.

THE USE OF INTELLIGENCE TESTS AS A BASIS OF SCHOOL ORGANIZATION AND INSTRUCTION

CHAPTER I

THE PLAN AND CONDUCT OF THE EXPERIMENT

The Problem. The experiment described in this bulletin was carried on in eight elementary schools in the city of Chicago. It was an attempt to answer the following question: What is the effect upon the efficiency of elementary schools of promoting and classifying pupils chiefly upon their mental ages and intelligence quotients as determined by group intelligence tests rather than according to the traditional method? It is recognized that this is really a double problem involving the question of a flexible system of promotion and classification upon any basis as compared with a non-flexible system, and also the question of using the results of group intelligence tests rather than some other basis for promotion and classification. The justification for combining these two questions is that the use of the results from group intelligence tests for the purposes mentioned above necessarily involves a flexible system and hence the two questions may be considered as one from the standpoint of practical school administration.

Definition of terms used in statement of problem. The "efficiency" of a school is the ratio of its output to the investment, or $\frac{\text{output}}{\text{investment}}$. The output or return upon the investment is measured in terms of the achievements of the pupils and their rates of progress through the school system. The investment, as the term is used above, includes not only what might strictly be called investment but also the factors which affect the manner in which the investment proper is used. The "promoting" of pupils refers to their advancement from one half-grade to another. The "classifying" of pupils refers to their placement in the fast, average and slow sections into which each half-grade in the experimental schools was divided. The word "chiefly" is used in the statement of the problem because the information derived from group intelligence tests was supplemented by other data. The "traditional method" refers to the method of placement used in the group of control schools. According to this method, promotion is determined by the pupil's final mark, which is usually a composite of the mark that he receives upon the final examination and that given by the teacher for his work during the term. In some cases the promotion indicated by the pupil's final mark is modified by the principal's opinion of his work or ability or by such considerations as chronological age, length of time already spent in the grade, number of pupils in the room, etc. This is the method which has been and still is the prevailing practice in the elementary schools of this country.

Scope of study. This study was confined to elementary schools having sixteen[1] or more teachers, which were organized in sixteen half-grades and in which pupils were promoted semi-annually. These schools were divided into an experimental and a control group of four each by Assistant Superintendent A. B. Wight. In making this selection, Mr. Wight endeavored to choose two groups of schools[2] in which the investment factors should be approximately equal at the beginning of the experiment.[3] Except in the plan of organization of the experimental schools, nothing was done to cause any change.

This investigation was rather strictly limited to the study of the effect upon the "efficiency" of certain elementary schools of promoting and classifying pupils chiefly according to the data derived from the use of group tests of intelligence. There was no consideration of the desirability of discovering and segregating for instructional purposes pupils of varying degrees of ability, except from the standpoint of their rates of progress and achievements in school. Neither was there any assumption that the plan used, which provided that pupils of different degrees of ability should complete the same course of study at different rates of progress, was superior to a plan providing that the different groups of pupils should cover different courses of study.

---

[1] In one of the control schools there were only fourteen teachers in charge of pupils who actually participated in the project.

[2] The experimental schools were the Armour, Franklin, Holden and Moseley. In the control group were the Alcott, Greene, Mark Sheridan and Webster.

[3] A more complete comparison of the investment factors in the two groups of schools may be found in the dissertation of the same title and by the same writer as this bulletin. This dissertation is on file in the library of the University of Illinois.

In this experiment the measurement of the achievements of the pupils was, with certain minor exceptions, limited to arithmetic and reading. As these are two of the most fundamental subjects studied in the elementary school, they were considered a fairly good measure of total achievement.

The general plan of the experiment. The experiment began in the autumn of 1920 and continued until the summer of 1922. As Table I shows, both intelligence and subject-matter tests were administered to the pupils of both groups of schools once each semester except that at the last testing only subject-matter tests were used. Also certain other data that seemed pertinent to the investigation were collected at each time of testing. After the first testing period the pupils of the experimental schools were promoted and classified upon the basis of the test results and the other data which had been obtained. After each of the later testing periods such adjustments were made as seemed advisable. No direct use was made of the test results or other data in the control schools. There was some opportunity for the teachers of this group of schools to make a more or less indirect use of the test results but they did not do so to a degree sufficient to affect the results of the experiment. By thus measuring the abilities and achievements of the pupils of the two groups of schools near the beginning and end of each semester the resulting data afforded a basis for comparing the effect of promoting and classifying pupils chiefly upon the results obtained from the use of group tests of intelligence with that of promoting pupils according to the traditional method.

The first tests were given in November, 1920, and the results used in promoting and classifying the pupils for the second semester of 1920-21. The next testing occurred in May, 1921, and furnished the basis for the placement of the pupils for the following September. All new entrants were tested in September and assigned to their grades and sections as soon thereafter as possible. A general testing occurred again in December and was followed by the placement of the pupils for the second semester of 1921-22. The final testing was in May, 1922. In addition to these general testing periods, small groups of absentees and new entrants were tested from time to time as seemed best.

The tests were in all cases given by the regular teachers who had been prepared for this work by a careful program of meetings with discussion. The teachers were also given very detailed directions. Rather extensive visiting by the writer while the tests were being administered showed that this program of preparation secured fairly uniform and correct procedure. Part of the scoring of the tests was done by the teachers and part by clerks in Assistant Superintendent Wight's office. The scoring was also checked sufficiently by the writer to warrant the belief that it was fairly accurate. Most of the errors which were found were so small that they had no effect upon the placement of pupils. In the tabulation of test and other data the positive and negative errors balanced each other so as to leave no sensible inaccuracy in the medians and other measures computed.

The data collected. Table I shows the intelligence and achievement tests used at each date of testing. The scores made upon these tests were translated into mental[4] or achievement[5] ages, as the case might be, and then further into intelligence[6] and achievement[7] quotients. The mental ages and intelligence quotients used

---

[4] Mental age is a term used to express the amount of intelligence possessed by an individual. The average score made upon an intelligence test by a large number of unselected children of any one given chronological age is said to be equal to a mental age of the given number of years. Thus, if on a given test the average score of six-year-olds is 25 points and that of seven-year-olds is 30 points, a score of 25 points may be transmuted into a mental age of six years and one of 30 points into one of seven years. It is abbreviated M.A.

[5] Achievement age is used to express the amount achieved by an individual on a subject-matter test. The average score made by the children of a single mental age is taken to equal an achievement age of the same number of years. Thus, if the average score of children of the mental age of ten years is 56 points, that score may be changed into an achievement age—abbreviated A.A.—of ten years.

[6] The intelligence quotient is the ratio of the mental age to the chronological age, or M.A. divided by C.A. It is conventionally carried to two places and written without the decimal point. Thus a child who has a mental age of ten years and is eight years old has an intelligence quotient of 10 divided by 8, or 125. It is abbreviated I.Q.

[7] The achievement quotient—abbreviated A.Q.—is the ratio of the achievement age to the mental age, or A.A. divided by M.A. It is written similarly to the I.Q. Thus a child whose achievement age is nine years and whose mental age is ten years has an achievement quotient of 9 divided by 10, or 90.

TABLE I. THE INTELLIGENCE AND ACHIEVEMENT TESTS USED IN THIS EXPERIMENT

Date of Testing	Grades	Intelligence Tests	Grades	Achievement Tests
Nov. 1920	IB-IIIB	Indiana University Primer Scale (Pressey Primer)	IA	Indiana University First Grade Reading Vocabulary Test, Form A
	IB-IIIB	Dearborn Group Tests of Intelligence	IIB-IIIB	Indiana University Scale of Attainment No. 1, Form A
	IIIA-VIIIA	National Intelligence Tests, Scale A, Form I		Monroe's Standardized Silent Reading Tests
			IIIA-VIB	Form I, Test 1
			VIA-VIIIA	Form I, Test 2
	IIIA-VIIIA	Illinois General Intelligence Scale, Form I		Monroe's General Survey Scale in Arithmetic
			IIIA-VIB	Form I, Scale 1
			VIA-VIIIA	Form I, Scale 2
May 1921	IB-IIIB	Indiana University Primer Scale		The same tests were used as in November, 1920, except that Form B of the two Indiana Tests and Form II of the Monroe Tests were used.
	IIIA-VIIIA	Illinois General Intelligence Scale, Form II		
Sept. 1921	IB	Kingsbury Primary Group Intelligence Scale, Form A		The same tests were used as in May, 1921
	IA-IIIB	Indiana University Primer Scale		
	IIIA-VIIIA	Illinois General Intelligence Scale, Form II		
Dec. 1921	IB-IIIB	Myers Mental Measure		The same tests were used as in November, 1920, except that Form III of the Monroe Tests was used.
	IIIA-VIIIA	Illinois General Intelligence Scale, Form I		
May 1922		None used.		The same tests were used as in November, 1920, except that Form B of the Indiana Vocabulary Test and Revised Form A of the Scale of Attainment No. 1 were used.
			VIB-VIIIA	Omnibus Test*

*See Appendix C.

at the first time of testing were based upon the average of the two mental tests given at that time.

The other items of information called for by the individual record cards used in this project were as follows: name, building, date of birth, sex, date of testing, school grade, chronological age, teacher's estimate, average school mark, attendance, and health mark.

The teacher's estimate was an opinion as to the general capacity of the pupil regardless of whether this capacity was actually displayed in regular school work or not. This estimate was expressed in terms of the following five marks and the teachers were instructed to make their distributions accord fairly closely with that given below:

S or 5=superior — 5 to 10 percent
E or 4=excellent — 20 percent
G or 3=good — 40 to 50 percent
F or 2=fair — 20 percent
P or 1=poor — 5 to 10 percent

The average school mark was the average of the pupil's marks upon the seven most important subjects of the course of study. It was based upon the grades on the monthly reports issued during the current semester previous to the date of testing and was expressed in terms of the same five marks that were used for the teachers' estimates.

Attendance was given as the percent of school days from the beginning of the semester to the date of testing during which the pupil was present. The health mark was the teacher's opinion of the general health of the pupil and was expressed in terms of the same five marks that were used for teachers' estimates and average school marks.

At the time of the first testing, the published norms and data for the transmutation of point scores upon the tests used into mental and achievement ages were in most cases based on a number of pupils not much larger, or even actually smaller, than the number taking the tests in this experiment. Hence it was decided that in the case of most of the tests used, the norms and tables for transmutation should be based upon the data obtained in this project.[8] The exceptions to this decision were the Illinois Examination, including the Illinois General Intelligence Scale and Monroe's Arithmetic and Reading Tests, and the Myers Mental Measure. These exceptions were made because in the case of the Illinois Examination scores from about fifty thousand pupils were available,[9] and in that of the Myers Mental Measure scores from about fifteen thousand pupils.[10]

---

[8] See complete dissertation for these transmutation tables and their derivation.

[9] Monroe, W. S. A Report of the Use of the Illinois Examination, Form 1, with 49,500 Pupils. Insert of School and Home Education, March, 1921. 8p.

[10] Myers, C. E. and G. C. Measuring Minds. New York; Newson, 1921, p. 23-4.

Principles of promotion and classification used at the first placement of the pupils. It was necessary to lay down certain principles which should be followed in the placement of the pupils, with the understanding that there would be need for exceptions in the cases of certain individuals. The inadvisability of following set rules too closely was due to several facts. Such procedure would result in entirely too great a change in the placement of some pupils. Moreover, the data secured from the tests and from other sources could not be relied upon as being absolutely accurate, and in some cases were so conflicting that disagreements between any detailed principles laid down were sure to occur. In view of these facts it must be understood that the principles enumerated below were not adhered to absolutely and that there were exceptions of many sorts that it is impracticable to list. The principles formulated for the first placement of pupils, which was for February, 1921, are given below.

I. The use of the data derived from the intelligence tests.

   1. The chief bases of placement were the mental ages and intelligence quotients. The mental ages were used to determine the half-grades in which the pupils should be placed, and the intelligence quotients to determine the sections, subject to such modifications as may be given in II.
   2. In general, the mental age norm for each half-grade was the median mental age of this half-grade group for the experimental schools. If the median mental age of a particular half-grade group in any one school was distinctly above or below the median of the four schools, a rough average of the two medians was used. This was done because it was considered desirable to make some progress toward reducing the range of ability within a given half-grade group for the experimental schools, but not to do so without regard to the ability actually found in the half-grades of the several schools as they were at the beginning of the experiment. Rather wide mental age limits were used for each half-grade group, with the expectation that as the experiment progressed they would be narrowed.

3. In general, demotion was recommended only when a pupil's mental age was at least two years below the median of the grade in which he was found, and extra promotion only when it was at least several months higher than the median of the grade in which extra promotion would place him. In no case was a pupil recommended for skipping more than two semesters' work nor for being demoted more than one.
4. Pupils whose intelligence quotients were above 110 were usually placed in the fast sections, those with I. Q.'s below 85 in the slow sections and the remainder[11] regularly composed the average sections. However, in many cases pupils' mental ages were several months above the medians of the half-grades in which normal promotion would place them, while their I. Q.'s were below 85. In such cases they were usually given normal promotion to the average section of the next half-grade. Similar exceptions were made in connection with other ranges of mental ages and intelligence quotients.
5. In some cases where the mental ages and intelligence quotients were rather low, it appeared probable that the pupils had either misunderstood directions upon one of the two intelligence tests or, had not, through some other cause, done themselves justice upon one of them. In such cases their scores on the other test were given more than half weight in determining their placement.

II. The use of the other data obtained.

1. Most of the other items recorded upon the individual record cards were given consideration. Low teachers' estimates and average school marks rarely prevented promotion in cases where the mental ages seemed to warrant it. In doubtful cases the question of whether

---

[11] Inasmuch as the intelligence quotients obtained in this project were derived from group intelligence tests they had a somewhat greater spread than those derived from individual tests. Therefore the percent of pupils with I.Q.'s from 85 to 110 was somewhat less than that usually found between 90 and 110 when individual tests are used.

a single or a double promotion should be given was, however, frequently decided by the teachers' estimates and average school marks.

2. On the other hand, even though their mental ages were low enough to merit demotion, very few pupils were failed whose teachers' estimates and average school marks were "good" or better, and not very many were failed if either one of the two was this high.
3. In making use of the teachers' estimates of capacity and the average school marks it was found that those of some teachers ran much higher than those of others in cases where the mental and achievement ages of the two groups of pupils showed little difference. That is, a teacher's estimate or school mark of "fair," for instance, given by one teacher might be fully equal to one of "good" given by another. In making use of these two items an allowance was made for this fact.
4. In cases of marked disagreement between the evidence afforded by the test data and that given by the teachers' estimates and school marks, the achievement test scores were frequently the deciding factor in placement.
5. In doubtful cases the teachers' estimates of health sometimes determined placement, but were not a major factor.

III. Pupils not classified.

1. Since all promoted VIIIB and VIIIA pupils would leave the schools concerned before the close of the experiment, no recommendations were made in the case of any pupils of these grades.
2. Pupils in open-air and ungraded rooms were recommended for promotion and classification according to the same principles used for the other pupils, but it was not expected that they would actually be placed in exact agreement with the recommendations.

An illustration of the application of these principles of promotion and classification. In order to illustrate the actual application of these principles, the following sample taken from the

TABLE II. SAMPLE OF THE PROMOTION LISTS MADE OUT FOR THE BEGINNING OF THE SECOND SEMESTER OF 1920–21

Pupil Number	M.A.	I.Q.	A.A.	T.E.*	School Mark	Health Mark	Placement of Pupil
1	6.6	73	8–8	1	1.8	3	IIIA slow
2	7.7	66	7–10	2	1.7	3	IVB slow
3	11.3	110	10–8	4	3.8	3	VB average
4	8.0	85	9–6	2	2.7	4	IVB fast
5	11.3	100	11–10	3	3.0	4	VB average
6	9.1	106	9–8	3	3.5	3	IVA average
7	9.2	95	8–8	3	3.1	3	IVA average
8	10.2	105	9–8	3	3.0	3	IVA average
9	10.6	78	11–10	3	3.0	4	IVA slow
10	8.8	71	10–0	2	2.0	1	IV B slow
11	4.7	40	7–4	1	1.0	3	IIIA slow
12	10.0	72	8–10	2	2.0	3	IVA slow
13	12.3	109	11–6	3	3.0	3	VB fast
14	10.4	91	11–2	3	3.0	4	IVA average
15	10.4	80	8–0	3	2.8	2	IVA slow
16	8.7	106	9–0	2	2.7	4	IVA average
17	10.2	69	13–0	3	3.0	4	VB average
18	9.2	106	8–6	3	3.2	3	IVA average
19	8.6	85	8–4	2	1.5	1	IVB slow
20	11.2	100	11–10	4	4.0	4	VB average

*Teacher's Estimate.

lists actually made out is given and discussed. These lists were later submitted to the principals and teachers, as has been mentioned previously, and any changes that seemed best were made.

Since the sample in Table II is a portion of the list for the pupils who were in the IVB grade during the first semester of 1920-21, the median mental and achievement ages[12] for the half-grade groups into which IVB pupils were likely to be placed are given below.

Grade	IIIA	IVB	IVA	VB
Mental Age	8.1	9.6	9.8	10.8
Achievement Age	7–4	9–1	9–2	10–6

The mental ages of Nos. 1 and 11 were so low that it was evident they should be demoted. This was corroborated by the low teachers' estimates and average school marks given them. Their I. Q.'s clearly indicated that they belonged in the slow section. Nos. 2, 10 and 19 had mental ages considerably below the IVB median and I. Q.'s of 85 or below. As their teachers' estimates and school marks were also fairly low they were kept in their grade and placed in the slow section. No. 12 had a mental age above the IVB median, but an I. Q. of only 72, so he was recommended for the IVA slow section. Although No. 4's mental age and I. Q. were low enough to indicate that he belonged in the IVB slow section his fairly good achievement age and his school mark of 2.7 resulted in his being placed in the fast section of that grade. This was done to prevent him from having to repeat work during the whole of the next semester and with the expectation that he would soon drop back into an average or slow section. The mental ages of Nos. 9 and 15 seemed to entitle them to extra promotion but as their teachers' estimates and school marks were only about average they received merely normal promotion into the IVA grade. Because of their low I. Q.'s they were placed in the slow section. In the case of the six pupils placed in the average section of IVA there was little doubt as to where they belonged except that No. 16 had a mental age almost a year below the IVA median. His rather high I. Q. and average school mark led to the decision not to prevent his advancement. Nos. 8 and 14 might have been considered for extra promotion had their teachers' estimates and school marks been higher. Nos. 3 and 20 were clearly entitled to extra promotion on the basis of all the data and No. 5 was only slightly less deserving. The I.Q. of the first would have caused his placement in the fast section but it happened there were not enough pupils in the school of similar ability to justify the formation of a fast section in grade VB. Therefore all three were placed in the average section of that grade. No. 17 was also given extra promotion. In his case a chronological age of almost 16 years and a high score on the achievement tests were potent clauses. For the same reason he was placed in the average rather than the slow section, although his I. Q. was only 69. No. 13 would probably have been given two semesters of extra promotion instead of one except for the fact that his teacher's estimate and school mark were only 3. As it was he was given one extra promotion and placed in the fast section of the grade.

---

[12] See Table III.

Supplementary principles of promotion and classification used at the second and third periods of placement of the pupils. At the second and third periods of placement—that is, for September, 1921, and February, 1922,—a majority of the pupils placed at the beginning of the second semester of 1920-21 received normal promotion into the next half-grade and remained in the corresponding section. Unless the new data clearly indicated that the pupil had been placed improperly in February, 1921, this course was followed. Certain additional principles were adopted to care for those pupils who seemed to have been improperly placed. These principles were as follows:

1. In the cases of a number of the pupils given extra promotion at the beginning of the previous semester, their school marks and achievement ages following this promotion did not appear to justify it. If, however, their mental ages as shown by the later testing were high enough to justify their retaining the extra promotion given and also receiving normal promotion at the later date, such promotion was usually given. This was done on the assumption that after skipping the work of one or more semesters it might require more than one semester for them to "find themselves."
2. Pupils previously promoted or placed in fast sections despite their low school marks were failed if their school marks still continued to be unsatisfactory.[13]
3. Many pupils who had received only a part of the extra promotion that they seemed to deserve in February, 1921,[14] were given further extra promotion, if their later scores justified so doing.
4. In cases where the test scores of pupils varied greatly from those made at the previous testing period or periods, and the other evidence did not agree with one score more than the other, the scores were roughly averaged to provide the basis for placement.

The final placement of the pupils. The writer made out his recommendations for placement, basing them upon the principles listed above, some two or three weeks before the end of the semester. The lists were then submitted to the principals and teachers concerned for their consideration, and finally put into effect. In two of the four schools the recommendations were discussed individually, but in the other two this was not done, as the principals of those schools wished to make as complete a change as possible from the traditional method of procedure. The changes made as a result of this consideration amounted to about one per cent of the total number of recommendations made. These changes were often due to the fact that a longer acquaintance with certain pupils caused the teachers to wish to revise the estimates of capacity or school marks which had been reported some time previously. Sometimes, however, the changes made represented a yielding on the part of the writer of his judgment, based largely upon the test results, to that of the principal or teacher, which was based upon the actual school work of the pupils and upon personal contact with them. In a few cases a change was made in order to place the pupil under a certain teacher so that he would be separated from a group of classmates.

---

[13] Most of these seemed to be cases of laziness and lack of study.

[14] These pupils had received only a part of their extra promotion in order to lessen the amount of work skipped at one time and thus make their advance easier.

In planning this whole experiment and in formulating and using the principles of promotion and classification those in charge of the experiment were guided by the desire to do a practicable piece of work. That is to say, they wished to use a procedure which the average school administrator or supervisor would be willing and able to make use of in his own school. It was partly because of this desire that more thoroughgoing changes were not made in the placement of the pupils, especially after the first period of testing. It is true that some public school superintendents have carried out considerably more radical plans of reclassification than the one used in this experiment but it was believed that a plan that might be followed by a more conservative educator would be more worth while.

CHAPTER II

CONDITIONS AT THE BEGINNING OF THE EXPERIMENT

Chronological age-grade placement in the two groups of schools. The chronological age-grade situation in November, 1920, showed that the retardation in the experimental schools was somewhat greater than that in the control schools. The median age of the pupils of the experimental schools averaged, grade for grade, two-tenths of a year more than that for the other group. In only two of the half-grades was it lower. The percents of pupils accelerated, normally placed, and retarded were 9, 18 and 73, respectively, in the experimental schools as compared with 10, 21 and 69 in the control schools. These figures are based upon the Chicago standard of normal progress, which is that a pupil should be from six to six and one-half years of age in grade IB, six and one-half to seven in grade IA and so on up. The average amount of retardation per pupil[1] was 1.14 years for the experimental schools and .96 year for the control schools. Assuming that pupils had entered the two groups of schools at the same average age, which the writer believes was the case, it is evident that the pupils in the control schools at the beginning of the experiment had made somewhat more rapid progress than had those in the experimental schools.

Mental age and school placement in the experimental and control schools. As may be seen from Table III, the median mental ages in all except three of the half-grades were higher in the control than in the experimental schools. The average difference was slightly over one-half year of mental age. This difference was found in spite of the fact just mentioned above that the pupils of the control schools were grade by grade about two-tenths of a year younger than those of the other group.

---

[1] The average amount of retardation was computed as follows: The number of pupils accelerated one-half year was multiplied by one-half, the number accelerated one year by one, and so on. The same process was carried out for those retarded and the sum found for each group of pupils. As the total number of years of retardation was greater than the total of acceleration, the latter was subtracted to give the net total of retardation. This was divided by the total number of pupils.

TABLE III. GRADE MEDIAN MENTAL AND ACHIEVEMENT AGES, INTELLIGENCE AND ACHIEVEMENT QUOTIENTS OF THE EXPERIMENTAL AND CONTROL SCHOOLS, NOVEMBER, 1920

Grade	Mental Ages		Intelligence Quotients		Achievement Ages		Achievement Quotients	
	Exp.	Cont.	Exp.	Cont.	Exp.	Cont.	Exp.	Cont.
IB	6.0	6.4	88	98				
IA	7.5	7.7	100	105	7–3	7–8	97	103
IIB	7.7	8.7	95	110	6–9	7–8	89	87
IIA	8.2	9.1	92	107	8–5	9–0	104	102
IIIB	9.0	8.6	100	96	10–1	9–10	113	113
IIIA	8.1*	8.3*	77*	84*	7–4*	7–10*	110*	110*
IVB	9.6	10.0	92	102	9–1	9–0	110	102
IVA	9.8	10.4	92	100	9–2	9–6	104	105
VB	10.8	10.7	93	98	10–6	10–7	106	108
VA	10.0	11.2	85	100	10–1	11–4	112	106
VIB	11.8	11.8	97	96	10–9	11–8	100	105
VIA	12.6	13.2	101	103	11–3	12–1	105	104
VIIB	12.5	14.4	99	112	11–6	13–0	105	100
VIIA	13.2	14.0	102	106	11–11	13–2	107	109
VIIIB	14.2	14.8	104	110	12–2	14–4	99	109
VIIIA	15.4	15.8	111	114	13–7	14–7	99	105
All	9.1	9.7	94	103	10–0	10–8	104	103

*The low M. A.'s and I. Q.'s found in grade IIIA were doubtless due to the fact that the Illinois General Intelligence Scale requires a degree of reading ability somewhat above that possessed by most IIIA pupils.

**In grades IIIA to VIIIA a composite achievement age was obtained by averaging the achievement ages upon Monroe's arithmetic scale and in comprehension and rate upon his reading test. In obtaining this average each of the three was given equal weight. The same procedure was followed in the case of the achievement quotients.

The inter-quartile ranges of the various grades were also computed. These showed an average range of two and one-tenth years for the experimental schools and two and two-tenths years for the control schools. Thus it appears that the grade groups in the experimental schools were slightly more homogeneous than those in the control schools. Similarly, a slight advantage was shown by the coefficients of correlation of mental age and grade placement. These were .84±.01[2] for the experimental schools and .82±.01 for the control schools.

The intelligence quotients of the two groups of schools. Probably the best basis of comparing the mentality of the pupils of the two groups of schools is that of their intelligence quotients. Table III shows the medians for the two groups of schools. In all grades except IIIB and VIB the median I. Q.'s of the experimental schools were lower than those of the control schools. The average difference was nine points, the medians for all grades combined being 94 and 103. This of course agrees with the fact just noted that the mental ages of the pupils of the experimental schools were lower, although their chronological ages were higher, than those of the other group.

---

[2] For convenience all probable errors smaller than .01 are given as .01

The extent to which the data derived from the tests afforded a true comparison of the quality of the pupil material of the two groups of schools depends upon the reliability of the tests used and the similarity of testing conditions in the two groups of schools. As is shown in Appendix B, the reliability of the intelligence tests was only fairly high, but there is no reason to think that the degree of reliability was different in the two groups of schools. Moreover, as has been stated in Chapter I, the writer's rather extensive visiting while the tests were being given and his examination of the test booklets after they had been scored afforded fairly reliable grounds for believing that there were no essential differences in the administration of the tests in the experimental and in the control schools.

The use of the control schools as a check group upon the experimental schools. If we assume that the difference in the amount and degree of intelligence found by the use of the intelligence tests was reliable, the question remains as to whether this difference was so great that the control schools could not be used as a valid check upon the experimental schools. A definite answer to this question cannot be given. Such data as are available concerning the mentality of pupils of different school systems appear to show that an average difference of about seven months of mental age or nine points I. Q. is not unusual. Probably the most extensive data available upon this point are those obtained from the use of the Illinois General Intelligence Scale.[3] This scale was given to the pupils of ten cities and nine counties in the autumn of 1920. It was found that the differences between the median mental ages of the various grades of the single cities and counties concerned and the general medians for the corresponding grades were four months or more in 50 percent of the cases. The largest difference was one year and three months. In terms of the I. Q. 50

---

[3] Monroe, W. S. A Report of the Use of the Illinois Examination, Form 1, with 49,500 Pupils. Insert of School and Home Education, March, 1921. 8p.

percent of the differences exceeded four points, the greatest being nineteen points. Differences as large as the average difference between the two groups of schools in this experiment were found in about one-sixth of the cases. Moreover, it must be remembered that differences between the individual members of a group and the group median are, on the average, much less than the differences between the individual members of the group. On the other hand, the differences in this experiment were based upon the average scores from two tests and therefore would probably tend to be smaller than those based upon a single score. This latter factor would not more than balance the one mentioned in the previous sentence, however, and probably would not even do that. Therefore the writer feels justified in the opinion that the difference in mentality found to exist between the two groups of schools was not so great but that the question referred to above can be answered affirmatively, provided that this difference was measured and taken account of in interpreting the results of the experiment.

The achievements of the two groups of schools. Table III also contains the median achievement ages for the various grades. It shows that the control schools were superior in all of the half-grades except IIIB and IVB. This average superiority was about eight months of achievement age, which is enough to indicate a decided superiority in pupil achievement on the part of the control schools.

The achievement quotients, however, are really more significant measures than are the achievement ages. It is evident from Table III that on the whole the relation of achievement to capacity, in so far as the tests used measured this relation, was practically the same in the two groups of schools. This would be inferred from a study of the mental and achievement ages. Such a comparison shows that the superiority of the control schools in achievement was just about the same as their superiority in intelligence. Thus from this standpoint the two groups of schools were capitalizing the capacities of their pupils almost equally in so far as the achievements measured were concerned.

The correlation of achievement with intelligence in the two groups of schools. It is a belief of many educators that the achievements of pupils should be as closely related to their capacities as possible and that the degree to which this relation holds is a measure of the success of the school in adapting its work to the individual pupils. The achievement quotient measures this from one standpoint, but it may also be measured by computing the correlation of achievement with intelligence. For all grades combined the coefficients of correlation between absolute achievement and intelligence scores was .68±.01 for the experimental schools and .60±.01 for the control schools. That is to say, the experimental schools were securing achievement more nearly in proportion to pupil capacity than were the control schools.

Another measure of the relation of achievement to intelligence may be obtained by computing the median achievement quotients for pupils of different levels of intelligence. This measure is based upon the assumption that the school should secure from all pupils the best work of which they are capable, and if it can not do this it should approach the standard as nearly for pupils of one level of intelligence as for those of another.

Table IV presents the median achievement quotients for the pupils of different levels of intelligence in the two groups of schools. A study of this table reveals the fact that in both groups of schools the inferior pupils were achieving more in relation to their capacity than were the superior pupils, but that this tendency was somewhat

TABLE IV. MEDIAN ACHIEVEMENT QUOTIENTS OF THE PUPILS OF DIFFERENT LEVELS OF INTELLIGENCE, NOVEMBER, 1920

Intelligence Quotient*	Achievement Quotients	
	Experimental	Control
150–59	104	98
140–	95	95
130–	102	98
120–	102	102
110–	101	101
100–	102	102
90–	104	104
80–	103	105
70–	107	110
60–	113	114
50–	109	125
All	105	104

*Only those levels of intelligence were included that had a sufficient number of cases to give fairly reliable medians.

less marked in the experimental schools. Further evidence to the same effect may be obtained from a comparison of the coefficients of correlation of the achievement and intelligence quotients. These were —.16±.01 for the experimental schools and —.28±.01 for the control schools.

Teachers' estimates of capacity, average school marks and estimates of health, in the two groups of schools. The pupil material of the two groups of schools may also be compared by means of the teachers' estimates, average school marks and health estimates. It is true that these measures are relatively subjective, but as there were almost one hundred teachers in each group of schools and as there was no apparent selection which would make one group of teachers more able to judge pupils than the other, these measures were probably fairly comparable for the two groups of schools. Taking the medians for all pupils, the teachers' estimates for the control group were two-tenths higher, the average school marks three-tenths higher, and the estimates of health two-tenths. Considering the three items together, the half-grade medians of the control schools were higher in about 50 percent of the cases, those of the experimental schools in only about 25 percent, and the two were equal in about 25 percent. The evidence afforded by these items is of value chiefly because it corroborates that obtained from the intelligence and achievement test results.

Summary. The differences found to exist between the experimental and the control schools in November, 1920, at the beginning of the experiment, were on the whole large enough not to be neglected as due to chance or as of no consequence, but were not large enough to invalidate the use of the two groups of schools in this experiment. When contrasted with the control schools the experimental schools exhibited the following differences:

1. .18 year greater retardation based on chronological age (1.14 years — .96 year)
2. .6 year lower median mental age (9.7 years — 9.1 years)
3. 9 points lower median I. Q. (103 — 94)

4. 8 months lower median achievement age in reading and arithmetic (10 years 8 months — 10 years)
5. 1 point higher achievement quotient in reading and arithmetic (104 — 103)
6. .08 higher correlation of achievement with intelligence. (.68 — .60)
7. .2 lower median teachers' estimate, school mark and estimate of health, averaged (3.3 — 3.1)

CHAPTER III

THE EFFICIENCY OF THE TWO GROUPS OF SCHOOLS AS MEASURED BY THE RATES OF PROGRESS OF THE PUPILS

In Chapter I "efficiency" was defined as the ratio of the output to the investment. The output to be measured was limited to the achievements of the pupils and their rates of progress. As was stated, the various factors constituting investment were all approximately constant except that of the mental abilities of the pupil material. Therefore, the "efficiency" of the experimental and the control schools might be measured in terms of the ratios of the achievements of the pupils and their rates of progress to their mental abilities. This chapter presents the data dealing with the rates of progress of the pupils, and the relation of these rates to their mental abilities.

The promotion and classification of the pupils for February, 1921. The information obtained from the testing in the four experimental schools in November, 1920, formed the chief basis for the placement of pupils for the succeeding semester, the second of 1920-21. This placement was made by the writer, following the principles of promotion and classification enumerated in Chapter I. The first half of Table V shows the percents of pupils in each grade of the experimental schools gaining or losing various amounts as a result of this placement. It is to be interpreted as follows, using grade IIB as an example: 7 percent of the pupils in grade IIB during the first semester of 1920-21 were demoted one semester, that is, were placed in IA; 26 percent were failed and remained in IIB; 64 percent received regular promotion of one semester into IIA and 2 percent received an extra promotion and thus entered IIIB. It will be seen that on the whole the placement of the pupils in the experimental schools involved many more demotions and failures than extra promotions and that the percent of the pupils given normal promotion was not as great as is usual in school systems. Only 58 percent of the pupils

TABLE V. PERCENTS OF THE PUPILS PROMOTED, FAILED OR DEMOTED THE GIVEN NUMBER OF SEMESTERS AT THE CLOSE OF THE FIRST SEMESTER OF 1920–21

First Semester Grade	Experimental Schools Semesters Gained or Lost					Control Schools Semesters Gained or Lost					
	−1*	0	+1	+2	+3	−1	0	+1	+2	+3	+4
IB		48	49	3	1		31	68	1		.2
IA	17	17	48	9	9		34	60	7		
IIB	7	26	64	2		2	9	80	9		
IIA	1	19	70	10		.4	16	83	.4		
IIIB	2	12	75	9	1		14	82	4	1	
IIIA	14	31	50	3	2		32	67	1		
IVB	4	19	57	13	8		10	82	7		
IVA	6	19	63	6	5	1	7	88	5		
VB	4	16	53	24	3		16	79	5		
VA	16	22	55	4	1		4	95	1		
VIB	10	28	43	15	4		11	87	2		
VIA	11	18	59	12			16	79	5		
VIIB	17	23	44	16			4	96	1		
VIIA	8	13	75	5			18	78	4		
All	6	26	58	8	2	.3	17	79	4	.03	.03

*−1 denotes one semester lost through demotion, O failure, +1 normal promotion, +2 one extra promotion, etc.

received normal promotion, 32 percent less, and 10 percent more. The average amount of promotion per pupil[1] was .74 semester. Probably the chief cause of the excess of demotions and failures and the low average promotion rate was the rather liberal promotion policy which had been pursued prior to the beginning of the project. Many decidedly inferior pupils were at the beginning of the experiment found to be almost up with normal pupils of the same chronological age, although they were unable to do satisfactory work as placed. Moreover, it had not been at all unusual to reward superior ability by allowing grades to be skipped.

The second half of Table V shows the changes made in the placement of the pupils in the control schools at this time. These changes were made by the teachers and principals according to the usual practice, which in Chapter I was called the "traditional method." A much larger percent of the pupils received normal promotion than in the experimental schools, but only about one-half as many were demoted or failed or given extra promotion. The average amount of promotion per pupil was .87 semester.

---

[1] The average amount of promotion per pupil was computed by finding the total number of semesters of promotion given, subtracting therefrom the total number of semesters of demotion and dividing by the total number of pupils concerned.

It is not fair, however, to compare the promotion rates of the two groups of schools directly according to the figures given above. The general assumption as to the promotion rate is that it should be one semester per semester for pupils of normal mentality who are properly classified and working to their full capacity with no hindering factors entering into the situation. For pupils whose mentalities are above or below normal and who are working under the same conditions as those mentioned for normal pupils the theoretical rates of progress are proportionately above or below one semester per semester. For example, a pupil with an I. Q. of 125 would be expected to advance one and one-fourth semesters per semester and one with an I. Q. of 80, four-fifths of a semester per semester. Thus to render the average promotion figures given in the preceding paragraphs strictly comparable each should be divided by the average or median I. Q. of the pupils concerned in order to bring both to the basis of what they would be for pupils of normal mentality, that is, pupils whose I. Q. is 100.

The measure of progress obtained by dividing the actual average rate of progress per pupil by the median intelligence quotient of the pupils contributing to this average will be called the "progress quotient." It will be used as the true measure of progress throughout this study. Making use of this measure we have as the "progress quotient" of the experimental schools .74 divided by .94, and for the control schools .87 divided by 1.03. Thus the "progress quotients" at this time were 79 for the experimental schools and 84 for the control schools. The true difference in the promotion rate at the beginning of the experiment is thus seen to have been only .05 rather than .13 semester. This difference cannot be attributed to the operation of the experimental plan of organization but rather to conditions in the two groups of schools previous to the beginning of the experiment and to the preparation necessary before the project could be begun.

The formation of the fast, average and slow sections. As a result of the placement in February, 1921, more pupils were placed in the slow sections and fewer in the fast sections than would usually be the case in most school systems. This was largely due to the liberal promotion policy that had been pursued before the experiment was begun, and to the fact that the pupil material of the experimental schools was rather distinctly inferior—median I. Q. 94, first quartile 80, third quartile 107. The fast sections included 14 percent of the total number of pupils placed, the average sections 41 percent, and the slow sections 45 percent.

The promotion and classification of the pupils for September, 1921. It was to be expected that after the experiment was under way a majority of the pupils in the experimental schools would make normal progress in the sections to which they had been assigned. The extent to which this expectation was fulfilled provided a measure of the efficiency of the previous placement. The supplementary principles of promotion and classification given in Chapter I suggest various reasons for the failure of many pupils to make such progress.

Table VI, which is similar to the first part of Table V, shows the gains and losses resulting from the placement of the pupils in the experimental schools for September, 1921. For example, in the average section of grade VIA 5 percent of the pupils were placed back in the VIA slow section and thus lost one-third of a semester; 17 percent placed in the VIIB slow section gained two-thirds of a semester; 63 percent placed in the VIIB average section gained one semester; 11 percent placed in the VIIB fast section gained one and one-half semesters and 4 percent placed in the VIIIA average section gained two semesters.

At this time 64 percent of the pupils of the experimental schools were advanced to the corresponding section of the next grade, but as such advancement meant only two-thirds of a semester for the slow pupils and one and one-half semesters for the fast, there were only 41 percent of the pupils who gained just one semester. In addition to the 64 percent mentioned, 26 percent were placed in the next grade but in a different section, so that in all 90 percent of the pupils of the experimental schools were advanced to the next grade. The average progress earned by the pupils in the slow sections was .65 semester, that earned by those in the average sections was .94 semester, and that by the members of the fast sections 1.38 semesters. For all the pupils the average was .88 semester. Dividing by the median I. Q.'s obtained from the Novem-

TABLE VI. PERCENTS OF THE PUPILS OF THE EXPERIMENTAL SCHOOLS GAINING OR LOSING THE GIVEN NUMBER OF SEMESTERS DURING THE SECOND SEMESTER OF 1920-21

Grade	Section	Semesters Gained or Lost										
		−1⅓*	−1	−⅓	0	+½	+⅔	+1	+1½	+1⅔	+2	+2½
IB	Slow				14	22	60	226				.4
	Average				21		57				.4	
IA	Slow			4	4		46	43	2			1
	Average		1		7		12	58	18		1	1
	Fast			1			4	22	65		1	7
IIB	Slow			6	5		62	23	3			
	Average			7	4		18	69			1	
	Fast				3		7	24	62			3
IIA	Slow			5			70	25				
	Average		1	2	2	1	7	73	10		1	2
	Fast							20	76			4
IIIB	Slow			6	1		67	24	1			
	Average	1				1	22	58	17		1	
	Fast						2	32	60		2	4
IIIA	Slow			14	2		68	14	1			
	Average			5	3		21	63	6	1	1	
	Fast						1	42	51		4	1
IVB	Slow			10			68	22				
	Average			3	2		23	61	5			7
	Fast						7	21	71			
IVA	Slow	2	1	7	6		60	23	1			
	Average	1		5			12	59	13		4	5
	Fast											
VB	Slow			6			63	30	2			
	Average			5			14	51	30			
	Fast							4	96			
VA	Slow	1		15			80	3	1			
	Average	1		6			17	52	24			
	Fast							16	84			
VIB	Slow			13			70	10	7			
	Average				2		25	54	16			2
	Fast							20	73			7
VIA	Slow			4			78	19				
	Average			5			17	63	11		4	
	Fast						5	20	75			
VIIB	Slow			2			76	20	2			
	Average			1			12	79	8			
	Fast								100			
VIIA	Slow				8		40	52				
	Average				1		4	76	19			
	Fast							6	94			
VIIIB	Slow			9			90	2				
	Average						4	96				
All	All	.2	.1	4	4	1	36	41	13	.1	1	1

*−1⅓ denotes a loss of one and one-third semesters, etc.

ber tests, the "progress quotients" were found to be 79, 92 and 118 for the slow, average and fast sections, respectively. For all the pupils of the experimental schools the quotient was 93.

Table VII shows the same data for the control schools as Table VI for the experimental. Of all the pupils in the control schools 80 percent were advanced just one semester. This percent is practically twice as large as that of the experimental schools, but is 10 smaller than the percent of pupils of those schools advanced to the same section of the next grade. Only 6 percent of the pupils of the control schools received extra promotion, as compared with the 15 percent in the experimental schools, but 14 percent were failed or demoted, as compared with only 8 percent in the latter group. The average progress in the control group was .92 semester. Dividing this by 1.02, a "progress quotient" of 90 was obtained. Thus, although the average progress of the pupils of the control schools was .04 semester greater, their "progress quotient" was three points smaller. To make the comparison upon a strictly valid basis, however, the effect of the pupils who left school and who entered school during the semester must be considered. Making the proper corrections for these pupils,[2] the average rate of progress

TABLE VII. PERCENTS OF THE PUPILS OF THE CONTROL SCHOOLS GAINING OR LOSING THE GIVEN NUMBER OF SEMESTERS DURING THE SECOND SEMESTER OF 1920–21

Grade	Semesters Gained or Lost							
	−2	−1	0	+1	+2	+3	+4	+5
IB			45	53	2			
IA		1	15	71	13	.4		
IIB			17	81	1			
IIA	1		10	88	1			
IIIB			12	87		.4	1	.4
IIIA		2	16	81	.4			
IVB			7	85	7	1		
IVA		2	9	88	1			
VB			11	87	2	1		
VA		1	4	83	12			
VIB			1	90	8	1		
VIA			9	75	16			
VIIB			1	90	9			
VIIA		.5	2	86	11			
VIIIB			4	96				
All	.1	1	13	80	6	.2	.1	.03

---

[2] See Appendix A.

of the pupils of the experimental schools was .02 semester less and their "progress quotient" four points larger than the corresponding figures for the pupils of the control schools. In other words, in so far as the progress of the pupils was concerned, the experimental schools were more efficient during the second semester of 1920-21 than were the control schools. The difference in the "progress quotient" was just about large enough to balance the difference at the beginning of the experiment.[3] Since this was the case, it cannot be assumed that the increased efficiency of the experimental schools in the matter of progress was necessarily due to the plan of organization used.

The relative size of the fast, average and slow sections. It was again deemed advisable to place many more pupils in the slow than in the fast sections. The facts that had made this necessary a semester earlier still exerted some influence upon the situation. The percent of the pupils placed in fast sections at this time was 15, and the remainder were equally divided between the average and the slow sections. Thus there was an increase of one percent in the number of pupils placed in the fast sections and also of those in the average sections over the percents for the previous semester.

The placement of the new entrants received by the experimental schools in September, 1921. The new entrants into the experimental schools in September, 1921, were tentatively placed in the average sections of the grades indicated by their previous school records, and later, after being tested, they were placed as the test results and the other data indicated. In the placement of pupils at this time no reclassification of the IB pupils was attempted, because the results obtained from the use of the Kingsbury Primary Group Intelligence Scale in that grade were so unsatisfactory that the writer deemed it unwise to make use of them. Table VIII shows that slightly over half of the pupils were not shifted at all, that 15 percent gained by the reclassification and 31 percent lost by it. The average change made amounted to a loss of .09 semester, or a promotion of .91 semester from the grades in which these pupils were the previous semester. As the median I. Q. of the new entrants was 85, their "progress quotient" was 107.

---

[3] See page 29.

TABLE VIII. PERCENTS OF THE NEW ENTRANTS INTO THE EXPERIMENTAL SCHOOLS GAINING OR LOSING THE GIVEN NUMBER OF SEMESTERS BY THEIR RECLASSIFICATION IN SEPTEMBER, 1921

Temporary Grade	Semesters Gained or Lost											
	−2⅓	−1⅓	−1	−⅓	0	+½	+⅔	+1	+1½	+1⅔	+2	+2½
IB*												
IA			27	13	53			7				
IIB		10	2	14	52	11	2	10				
IIA		8		35	42	8		4	4			
IIIB		7		7	51	21		14				
IIIA				23	77							
IVB		10	2	17	58	6		6	2			
IVA				38	52	5					5	
VB		7		25	53	2		14				
VA		19		21	47	8	4	2				
VIB		9		14	59			9	5			5
VIA		17	13	17	43			4	4			
VIIB	3			12	64	15		3		3		
VIIA		29	7	7	50				7			
VIIIB		12	31	15	42							
VIIIA					100							
All	.2	9	4	18	53	7	.2	6	1	.2	.2	.2

*As is explained in the text, the pupils in grade 1B were not reclassified at this time.

Only 9 percent of the new entrants at this time were placed in the fast sections. The average sections received 63 percent and the slow sections 28 percent. Combining the new entrants with the pupils who had been tested in the previous May the percent in the fast sections was 15, that in the average 45, and that in the slow 41.

The promotion and classification of the pupils for February, 1922. The data obtained in December, 1921, were used to determine the placement of the pupils for the second semester of the school year. Table IX, which is similar to Table VI, gives the percents of the pupils of the experimental schools gaining or losing various amounts during the first semester of 1921-22. There were 48 percent of the pupils advanced just one semester, as compared with 41 percent during the previous semester; 35 percent made less than one semester's progress and 17 percent made more, as compared with 44 and 15 percent previously. The percent of the pupils advanced to the corresponding section of the next higher grade was 79, whereas only 64 percent were so advanced a semester previously. The percent placed in some section of the next grade

TABLE IX. PERCENTS OF THE PUPILS OF THE EXPERIMENTAL SCHOOLS GAINING OR LOSING THE GIVEN NUMBER OF SEMESTERS DURING THE FIRST SEMESTER OF 1921-22

Grade	Section	Semesters Gained or Lost																
		−1⅔	−1	−⅔	−½	−⅓	0	+⅓	+½	+⅔	+⅚	+1	+1⅓	+1½	+1⅔	+2	+2⅓	+2½
IB	Slow						7			61		27				5		
	Average						24			12		62				2		
IA	Slow						14	8		65			12					
	Average		1			3	6	7				76				7		
	Fast						2							56		41		
IIB	Slow					12		9		58			16		1		4	
	Average			2			4	4				86	2			2		
	Fast								3		1			95				1
IIA	Slow					12		8		73			8					
	Average	1		4				4				89				2		
	Fast								10		5			85				
IIIB	Slow					6				80			14					
	Average			1			5	3				84				7		
	Fast								2					98				
IIIA	Slow					6		2		82			10					
	Average	1		1			7	8				83						
	Fast								3					97				
IVB	Slow					4		2		79			16					
	Average			3			3	7				84				3		
	Fast										17			83				
IVA	Slow					1		1		76			22					
	Average			4				3				78	5			10		
	Fast													100				
VB	Slow					4				78			18					
	Average		1	5				4				85				5		
	Fast													100				
VA	Slow									92			7				1	
	Average			1			6	3				85	1			4		
	Fast								2					98				
VIB	Slow					1				63			36					
	Average						6	7				87						
	Fast				3									97				
VIA	Slow									59			37				4	
	Average	1	2	5			9	5				75				2		
	Fast													100				
VIIB	Slow					3				87			10					
	Average			1				4				94				1		
	Fast													100				
VIIA	Slow									74			26					
	Average			1			2					96						
	Fast													100				
VIIIB	Slow									83							17	
	Average			4			2	7				80				8		
	Fast													100				
VIIIA	Slow									100								
	Average							2				98						
All	All	.1	.1	1	.02	1	6	3	.2	23	.2	48	5	10	.02	2	.2	.02

was almost the same as before, being 89. The average progress per pupil was .72 semester for those in the slow sections, .87 semester for those in the average and 1.49 semesters for the members of the fast sections. The respective "progress quotients" were 84, 84 and 121. For all the pupils in the experimental schools the average progress was .90 semester and the "progress quotient" 91. The corresponding figures for the second semester of 1920-21 were .88 semester and 93, so it is apparent that the average progress was slightly greater and the "progress quotient" slightly less during the second semester of the experiment than during the first.

Table X, which is similar to Table VII, shows the gains and losses of the pupils of the control schools according to their placement at this time. A comparison of these data with those for the experimental schools shows that, as before, a larger percent of the pupils of the control schools received normal promotion. The difference, however, was not quite as great as the previous semester, the figures for this time being 82 and 48 percent as compared with 80 and 41 percent. The percent of the pupils receiving extra promotion in the control schools was only half as large as it had been a semester earlier, whereas in the experimental schools the corre-

TABLE X. PERCENTS OF THE PUPILS OF THE CONTROL SCHOOLS GAINING OR LOSING THE GIVEN NUMBER OF SEMESTERS DURING THE FIRST SEMESTER OF 1921–22

Grade	Semesters Gained or Lost						
	−3	−2	−1	0	+1	+2	+3
IB				33	64	2	.2
IA				19	80	.5	
IIB				14	82	3	
IIA				17	83	1	
IIIB	.3		.3	6	91	3	
IIIA				20	80		
IVB				10	83	7	
IVA				6	92	2	
VB				10	84	4	2
VA			1	6	91	.3	
VIB		1		11	85	4	
VIA	1			13	82	4	1
VIIB				11	88	1	
VIIA			1	8	79	13	
VIIIB				9	91		
VIIIA	1		2	9	89		
All	.1	.03	.1	14	82	3	.2

sponding percent was slightly larger. The percent of failures and demotions in each group was practically the same as before. The average progress per pupil in the control schools was .88 semester and the "progress quotient" was likewise 88. Hence the average progress was .02 semester greater in the case of the experimental schools and the "progress quotient" three points greater. Had it not been for the new entrants and eliminees, the difference in average progress would have been .01 semester greater. Thus it can be said for the second semester of the experiment, as for the first, that in so far as the progress of the pupils was concerned, the experimental schools were somewhat more efficient than were the control schools.

The classification into fast, average and slow sections for the second semester of 1921-22. The percents of the pupils placed in the sections at this time differed rather markedly from those for previous semesters. The percent placed in the fast sections showed only a slight decrease, but that in the slow sections decreased about one-third. The percents were 13 in the fast sections, 57 in the average and 30 in the slow sections. These figures give evidence that as the experiment progressed it was possible to place pupils more nearly as would be expected from theoretical considerations.

The promotion and classification of the pupils for September, 1922. After the testing in May, 1922, which was the last during the experiment, the pupils of the experimental schools were placed for the first semester of 1922-23. Table XI, which is similar to Tables VI and IX, shows the gains and losses of the pupils of the experimental schools during the second semester of 1921-22. There were 55 percent of the pupils who gained just one semester as compared with 48 percent during the previous semester, 28 percent who made less than one semester's progress as compared with 35 percent, and 17 percent who made more, the same as the previous semester. Only 58 percent of the pupils were advanced to the corresponding section of the next higher grade as compared with 79 percent a semester previously. The percent placed in some section of the next grade was 89, just the same as it had been. The average progress per pupil was .62 semester for those in the slow section, .98 semester for those in the average, and 1.39 for

TABLE XI. PERCENTS OF THE PUPILS OF THE EXPERIMENTAL SCHOOLS GAINING OR LOSING THE GIVEN NUMBER OF SEMESTERS DURING THE SECOND SEMESTER OF 1921-22

Grade	Section	Semesters Gained or Lost													
		−1⅓	−1	−⅔	−⅓	0	+⅓	+½	+⅔	+1	+1½	+1⅔	+2	+2⅙	+2½
IB	Slow					37			46	17					
	Average					12			22	66					
IA	Slow			1		30	33			36					
	Average					20				69	11		.3		
	Fast					3			5	21	66		5		
IIB	Slow				10				90						
	Average					9				75	4		12		
	Fast									21	79				
IIA	Slow				5				94			2			
	Average					5				81	12		2		
	Fast									4	83		13		
IIIB	Slow				19				81						
	Average					12				81	7				
	Fast									16	65		18		
IIIA	Slow				11				89						
	Average					5				79	14		2		
	Fast										95				5
IVB	Slow				6				94						
	Average					5				77	18				
	Fast									19	78		4		
IVA	Slow				6				94						
	Average		1			6				82	8		3		
	Fast										100				
VB	Slow	1			3				96						
	Average					2				93	4		1		
	Fast										100				
VA	Slow								100						
	Average					3				86	7		3		
	Fast														
VIB	Slow								99					1	
	Average					2				85	6		6		
	Fast									8	92				
VIA	Slow								100						
	Average					5		6		83	6				
	Fast									18	82				
VIIB	Slow				3				26			71			
	Average					3				92	2		2		
	Fast									31	63		6		
VIIA	Slow				6				94						
	Average									89	8		3		
	Fast										100				
VIIIB	Slow				3				97						
	Average					4				96					
	Fast									38			63		
VIIIA	Slow									100					
	Average					4				96					
All	All	.03	.03	.03	1	6	1	.2	20	55	14	1	2	.03	.03

members of the fast sections. The respective "progress quotients" were 76, 98 and 114. For all pupils of the experimental schools the average progress was .94 semester and the "progress quotient" 97. These figures show both greater actual progress and greater progress relative to ability than was made during either of the previous semesters.

Table XII, which is similar to Tables VII and X, shows the gains and losses of the pupils of the control schools for this semester. Again more pupils of the control schools received normal promotion than was the case in the experimental schools, the difference, however, being smaller than it was in either of the previous semesters. The percent of the pupils receiving extra promotion was only one-third as large as in February, 1922, whereas in the experimental schools it was the same. The percent of failures and demotions in the control schools was slightly less than in February, the decrease being in about the same ratio as that in the experimental schools. The average progress per pupil in the control schools was .89 semester and the "progress quotient" 86. Thus the average progress was .05 semester greater in the case of the experimental schools and the "progress quotient" eleven points greater.

TABLE XII. PERCENTS OF THE PUPILS OF THE CONTROL SCHOOLS GAINING OR LOSING THE GIVEN NUMBER OF SEMESTERS DURING THE SECOND SEMESTER OF 1921–22

Grade	Semesters Gained or Lost				
	−1	0	+1	+2	+3
IB		8	69	23	
IA	.4	18	81	1	
IIB		16	83	2	
IIA		10	90		
IIIB		8	91	1	
IIIA		24	75	1	
IVB		8	90	2	
IVA		10	90		
VB		13	86	1	1
VA		6	94		
VIB		8	92		
VIA		7	92	1	
VIIB	3	9	85	1	1
VIIA		16	83	1	
VIIIB		17	83		
VIIIA		5	95		
All	.2	12	87	1	.1

The new entrants and eliminees did not affect these differences.

The classification into fast, average and slow sections for the first semester of 1922-23. The percents of the pupils placed in the sections for September, 1922, again differed considerably from those for previous semesters. The percent in the fast sections was practically the same, 14, but that in the average sections rose to 68 and that in the slow sections dropped to 18. Thus the tendency already noted for the fast and slow sections to approximate each other in size was continued as the experiment progressed longer. Probably the distribution at this time was about what it should be, as there will always be more pupils belonging in slow sections because of not realizing their highest possible achievement than there will be pupils belonging in fast sections because of doing more than should be expected of them.

Summary. As a result of the placement of the pupils at the beginning of the experiment the "progress quotient" for the experimental schools was, at that time, five points smaller than that for the control group. During the course of the experiment this situation was reversed. Averaging the "progress quotients" for the three semesters, those for the experimental group were the larger by about six points. Thus the net result of the experimental plan of organization in so far as progress was concerned was favorable. The greater degree of efficiency of the experimental schools seems to have been due to the operation of this plan.

CHAPTER IV

THE EFFICIENCY OF THE TWO GROUPS OF SCHOOLS AS MEASURED BY THE ACHIEVEMENTS OF THE PUPILS

In Chapter III, one of the two factors which were taken as constituting output has been discussed. In this chapter the other factor, that of achievement, will be considered both absolutely and in its relation to capacity. It has been shown in the preceding chapter that there was an increase in the "progress quotient" of the experimental schools as compared with that of the control schools. Therefore if a study of the achievements of the pupils of the two groups of schools shows that those of the experimental schools were either equal to or greater than those of the control schools, it may be said that the experimental schools were more efficient than the other group during this experiment.

The gains in absolute achievement during the second semester of 1920-21. Table XIII gives the median achievement ages of the grades and sections of the two groups of schools at the beginning and end of the second semester of 1920-21. A comparison of the first and fourth columns shows that at the beginning of this semester the median achievement age of the control schools was four months greater than that of the experimental schools, while at the end of this semester the median ages were the same. In other words, the gain in achievement age on the part of the experimental schools was four months more than that of the control schools. The cause of this increase cannot be stated with certainty. There are at least two explanations that may account for it. One of these is that it resulted from the same causes which accounted for a similar increase in the mental ages. The increase in the median mental age of the pupils of the experimental schools was five-tenths of a year greater during this semester than was that of the control schools. In the opinion of the writer the most potent cause of the greater increases in both mental and achievement ages on the part of the experimental schools was the fact that both the teachers and the pupils of those schools felt a very high degree of

TABLE XIII. MEDIAN GRADE AND SECTION ACHIEVEMENT AGES OF THE PUPILS OF THE TWO GROUPS OF SCHOOLS AT THE BEGINNING AND END OF THE SECOND SEMESTER OF 1920–21

Grade	Section	Experimental			Control		
		Beginning	End	Gain*	Beginning	End	Gain
IA	Slow	6–4	10–5	49	6–2	10–0	46
	Average	6–8	10–4	44			
	Fast	5–6	10–10	64			
IIB	Slow	6–4	8–8	28	7–11	8–0	1
	Average	6–5	9–10	41			
	Fast	8–6	8–4	–2			
IIA	Slow	6–7	9–5	34	7–10	9–7	21
	Average	7–7	9–10	27			
	Fast	8–1	9–6	17			
IIIB	Slow	8–2	10–10	32	9–0	10–9	21
	Average	8–11	10–7	20			
	Fast	8–8	11–1	29			
IIIA	Slow	9–1	9–1	0	9–8	10–1	5
	Average	10–1	10–1	0			
	Fast	9–10	9–8	–2			
IVB	Slow	8–11	10–8	21	9–1	10–10	21
	Average	10–0	10–7	7			
	Fast	9–11	12–6	31			
IVA	Slow	9–5	11–7	26	10–0	10–0	0
	Average	10–7	14–0	41			
	Fast	....	....	...			
VB	Slow	9–8	12–1	29	10–7	12–2	19
	Average	10–10	13–4	30			
	Fast	10–10	13–2	28			
VA	Slow	10–4	11–4	12	11–5	13–4	23
	Average	11–5	13–6	25			
	Fast	13–0	14–7	19			
VIB	Slow	10–0	12–1	25	12–0	13–8	20
	Average	11–10	12–10	12			
	Fast	13–6	15–1	19			
VIA	Slow	11–11	14–5	30	12–1	13–10	21
	Average	12–8	15–1	29			
	Fast	15–0	15–0	0			
VIIB	Slow	12–5	15–5	36	14–2	14–8	6
	Average	12–5	16–6	49			
	Fast	15–4	17–0	20			
VIIA	Slow	12–7	14–11	28	15–0	16–1	13
	Average	13–6	17–8	50			
	Fast	16–1	18–5	28			
VIIIB	Slow	11–11	13–6	19	15–1	16–1	12
	Average	14–0	16–7	31			
All	All	10–0	11–6	18	10–4	11–6	14

*The gains are given in terms of months.

TABLE XIV. MEDIAN GRADE AND SECTION ACHIEVEMENT QUOTIENTS OF THE TWO GROUPS OF SCHOOLS AT THE BEGINNING AND END OF THE SECOND SEMESTER OF 1920–21

Grade	Section	Experimental			Control		
		Beginning	End	Gain	Beginning	End	Gain
IA	Slow	103	116	13	91	117	26
	Average	88	118	30			
	Fast	65	128	63			
IIB	Slow	93	100	7	99	95	−4
	Average	86	109	23			
	Fast	103	105	2			
IIA	Slow	89	112	23	90	111	21
	Average	92	111	19			
	Fast	88	110	22			
IIIB	Slow	111	120	9	103	118	15
	Average	106	111	5			
	Fast	103	112	9			
IIIA	Slow	114	104	−10	115	112	−3
	Average	114	115	1			
	Fast	108	104	−4			
IVB	Slow	111	106	−5	110	112	2
	Average	106	106	−0			
	Fast	106	108	2			
IVA	Slow	107	110	3	103	103	0
	Average	106	118	12			
	Fast	...	...	...			
VB	Slow	109	112	3	105	110	5
	Average	108	109	1			
	Fast	98	108	10			
VA	Slow	106	108	2	106	112	6
	Average	110	116	6			
	Fast	105	118	13			
VIB	Slow	102	110	8	109	114	5
	Average	104	108	4			
	Fast	110	106	−4			
VIA	Slow	115	119	4	105	115	10
	Average	108	118	10			
	Fast	115	121	6			
VIIB	Slow	108	120	12	103	108	5
	Average	105	122	17			
	Fast	105	117	12			
VIIA	Slow	108	106	−2	101	113	12
	Average	103	119	16			
	Fast	105	122	17			
VIIIB	Slow	99	104	5	105	111	6
	Average	93	113	20			
All	All	106	112	6	104	112	8

interest in the results of the tests because they knew that placement was largely dependent upon these results. On the other hand, the teachers and pupils of the control schools knew that no direct use would be made of the test results, hence naturally took less interest in the testing. The other explanation is that the experimental plan of organization caused the increase. Inasmuch as there is generally a fairly high correlation between the scores made on intelligence tests and those on subject-matter tests, especially in the case of verbal intelligence and reading tests, the writer believes that the first explanation is the true one or at least more nearly so than the latter. It is not unlikely that both had a part in causing the relative increase.

The achievement quotients at the beginning and end of the second semester of 1920-21. Table XIV shows that at the beginning of this semester the median achievement quotient of the experimental schools was two points higher than that of the control schools. We have seen that both the intelligence and the achievement scores made at the end of the semester showed a greater increase in the case of the experimental schools than in that of the other group, but that the increase in intelligence on the part of the experimental schools was slightly greater than that in achievement. Moreover, both groups showed greater increases in achievement than in intelligence. Therefore we expect to find, as we do, that the median achievement quotients of both groups of schools increased during this semester, and that the increase in the case of the control schools was slightly greater. This difference was two points. Thus at the end of the semester the two medians were the same. The general import of this evidence is that in so far as achievement was concerned there was a slight relative increase in the efficiency of the control schools.

The correlation of intelligence and achievement at the beginning and end of the second semester of 1920-21. Although the achievement quotient measures the relation of intelligence and achievement in one way, this relation may also be measured by means of the coefficient of correlation. The following table shows the coefficients that were found at the beginning and end of the semester by correlating the mental and achievement ages for all grades combined:

Experimental			Control		
Beginning	End	Loss	Beginning	End	Loss
.68±.01	.56±.01	.12	.60±.01	.51±.01	.09

It is evident that in both groups there was a decrease in the correlation of achievement with intelligence as measured by the tests used. This decrease was slightly greater in the case of the experimental schools, but the difference was not great enough to be significant. It may be that this decrease was due to a lessening of the degree to which instruction was adapted to the capacities of the pupils. It is likely, however, that much, if not all, of the decrease was caused by the fact that the mental ages calculated at the beginning of the semester were based upon the average scores made on two intelligence tests and hence were more reliable than those obtained at the end of the semester, which were based upon only one test score. This lower degree of reliability would naturally tend to reduce the correlation between the mental and the achievement ages.

The relation of the intelligence and achievement quotients at the beginning and end of the second semester of 1920-21. The median achievement quotients of the groups of different levels of intelligence as determined by the intelligence quotients were computed for the end of the semester as they had been at its beginning. Table XV presents a comparison of those found at the two periods. The achievement quotients of the different groups at the end of the semester showed that in the experimental schools instruction was adapted about equally well to the pupils of different

TABLE XV. MEDIAN ACHIEVEMENT QUOTIENTS OF THE PUPILS OF DIFFERENT LEVELS OF INTELLIGENCE AT THE BEGINNING AND END OF THE SECOND SEMESTER OF 1920–21

Intelligence Quotient	Achievement Quotients			
	Experimental		Control	
	Beginning	End	Beginning	End
150–59	104	117	98	103
140–	95	122	95	113
130–	102	113	98	107
120–	102	113	102	109
110–	101	112	101	109
100–	102	113	102	113
90–	104	111	104	113
80–	103	113	105	114
70–	107	115	110	117
60–	113	116	114	120
50–	109	128	125	120
All	105	112	104	112

levels of intelligence, whereas in the control schools the previous well-marked tendency to adapt instruction more nearly to the capacities of the inferior pupils remained. As this tendency was present in both groups of schools at the beginning of the semester it is evident that there was a relative improvement in the degree to which the instruction in the experimental schools was adapted to pupils of one level of intelligence as well as to those of another.

The coefficients of correlation of the intelligence and achievement quotients were also found and compared with those for the beginning of the semester. The following table presents this comparison:

Experimental			Control		
Beginning	End	Gain	Beginning	End	Gain
—.16±.01	—.12±.01	.04	—.28±.01	—.25±.01	.03

This comparison shows that at the close of the semester the negative correlations between the intelligence and achievement quotients were slightly smaller in the cases of both of the groups of schools. The difference in the gains was so small that it has no significance. This fact shows that the instruction given in the control schools was still somewhat less equally suited to pupils of all levels of intelligence than was that of the experimental schools. In the main this corroborates the evidence presented in the preceding paragraph.

The gains in absolute achievement during the first semester of 1921-22. Table XVI, which contains data corresponding to the third and sixth columns of Tables XIII and XIV, shows that the increase in absolute achievement during this semester was seven months of achievement age in the experimental schools and four months in the control schools. As during this same semester the median mental age of the experimental schools did not increase as much as did that of the control schools it seems fair to attribute the greater gain in achievement to an increase in the efficiency of the experimental schools. It was shown in the preceding chapter that during this same semester the "progress quotient" of this group of schools was greater than that of the control group, therefore this increase in efficiency in so far as achievement is concerned cannot be attributed to a slowing up of the progress of the pupils.

TABLE XVI. GRADE AND SECTION GAINS AND LOSSES IN ACHIEVEMENT AGES AND QUOTIENTS OF THE TWO GROUPS OF SCHOOLS FROM THE BEGINNING TO THE END OF THE FIRST SEMESTER OF 1921–22

Grade	Section	Achievement Ages		Achievement Quotients	
		Experimental	Control	Experimental	Control
IA	Slow	16	−2	12	6
	Average	23		7	
	Fast	8		−19	
IIB	Slow	−20	−18	−11	−31
	Average	−14		−19	
	Fast	−11		−13	
IIA	Slow	11	6	18	−21
	Average	7		−2	
	Fast	5		10	
IIIB	Slow	8	12	−1	−32
	Average	17		−11	
	Fast	16		−8	
IIIA	Slow	−15	−16	−11	−12
	Average	−17		−7	
	Fast	−10		−10	
IVB	Slow	10	4	−4	−9
	Average	15		−4	
	Fast	21		−5	
IVA	Slow	23	−7	1	−18
	Average	22		1	
	Fast	67		12	
VB	Slow	20	17	−11	−1
	Average	5		−10	
	Fast	−2		−25	
VA	Slow	12	18	−7	2
	Average	−6		−15	
	Fast	6		−9	
VIB	Slow	17	8	−8	−8
	Average	19		−3	
	Fast	−49		−12	
VIA	Slow	14	8	−13	−5
	Average	2		−11	
	Fast	−29		−2	
VIIB	Slow	19	0	−9	−12
	Average	18		−10	
	Fast	35		2	
VIIA	Slow	5	3	−13	−5
	Average	10		−3	
	Fast	27		−3	
VIIIB	Slow	−19	1	−15	−9
	Average	−8		−13	
	Fast	−15		−18	
VIIIA	Slow	7	−3	−7	−17
	Average	12		3	
All	All	7	4	−8	−13

The achievement quotients at the beginning and end of the first semester of 1921-22. The median achievement quotients of both groups of schools were smaller at the end of this semester than they were at its beginning. In other words, the average increase in the scores made upon the intelligence tests was considerably greater than that in those upon the achievement tests. This would seem to point to the fact that the practice effect upon the intelligence tests was greater than that upon the others. As Table XVI shows, the decrease in the median achievement quotient of the experimental schools was eight points, whereas that in the control schools was thirteen points. Thus the loss of the experimental schools was five points less than that of the other group, or, in other words, their relative gain was that large. This supports the conclusion reached above from the study of the mental and achievement ages, that the efficiency of the experimental schools during the semester was greater than that of the control schools.

The correlation of achievement and intelligence at the beginning and end of the first semester of 1921-22. The following table compares the coefficients of correlation found at the end of the semester with those at the beginning:

Experimental			Control		
Beginning	End	Loss	Beginning	End	Loss
.56±.01	.54±.01	.02	.51±.01	.51±.01	.00

Judging from these coefficients, it seems that there was practically no change in the relation of achievement to intelligence during this semester. The slight decrease of .02 on the part of the experimental schools was too small to have any significance.

The relation of the intelligence and achievement quotients at the beginning and the end of the first semester of 1921-22. Table XVII shows the same facts for this semester as Table XV for the previous semester. At the end of this semester there was a rather definite decrease in the achievement quotient medians of the experimental schools from the duller to the brighter pupils. This tendency was even more marked in the control schools. Inasmuch as at the beginning of the semester this tendency was not noticeable in the experimental schools but was present in the control schools the figures for the end of the semester indicate that relatively the condition which they measure became worse in the

TABLE XVII. MEDIAN ACHIEVEMENT QUOTIENTS OF THE PUPILS OF DIFFERENT LEVELS OF INTELLIGENCE AT THE BEGINNING AND END OF THE FIRST SEMESTER OF 1921–22

Intelligence Quotient	Achievement Quotients			
	Experimental		Control	
	Beginning	End	Beginning	End
150–59	117	98	103	85
140–	122	96	113	88
130–	113	102	107	95
120–	113	100	109	99
110–	112	102	109	104
100–	113	106	113	105
90–	111	106	113	105
80–	113	110	114	107
70–	115	111	117	113
60–	116	114	120	119
50–	128	125	120	132
All	112	105	112	104

experimental schools. That is to say, during this semester there was a relative loss in the degree to which instruction was equally well adapted to pupils of all levels of intelligence in the experimental schools. The writer is unable to suggest any probable explanation of this fact.

A comparison of the coefficients of correlation of the intelligence and achievement quotients at the end of the semester with those at the beginning supports the conclusion given above. These coefficients were as follows:

Experimental			Control		
Beginning	End	Loss	Beginning	End	Loss
—.12±.01	—.36±.01	.24	—.25±.01	—.39±.01	.14

This comparison shows that the correlation between the intelligence and achievement quotients became considerably greater, negatively, during the semester. The change was much larger in the experimental schools. This fact emphasizes the conclusions presented in the last two paragraphs to the effect that there was a relative decrease in the degree to which the experimental schools capitalized the capacities of their pupils into achievement. This decrease is even more definitely shown by these coefficients than by the data given previously.

The gains in absolute achievement during the second semester of 1921-22. Table XVIII, which is similar to Table XVI,

TABLE XVIII. GRADE AND SECTION GAINS AND LOSSES IN ACHIEVEMENT AGES AND QUOTIENTS OF THE TWO GROUPS OF SCHOOLS FROM THE BEGINNING TO THE END OF THE SECOND SEMESTER OF 1921–22

Grade	Section	Achievement Ages		Achievement Quotients	
		Experimental	Control	Experimental	Control
IA	Slow	35	18	25	4
	Average	40		45	
	Fast	47		16	
IIB	Slow	3	0	7	−4
	Average	7		12	
	Fast	−7		5	
IIA	Slow	11	15	7	17
	Average	10		10	
	Fast	6		25	
IIIB	Slow	1	16	−3	19
	Average	7		2	
	Fast	7		10	
IIIA	Slow	−16	−4	−6	13
	Average	−15		−9	
	Fast	−2		16	
IVB	Slow	6	16	1	12
	Average	4		9	
	Fast	15		3	
IVA	Slow	5	19	7	11
	Average	13		9	
	Fast	12		20	
VB	Slow	0	12	8	11
	Average	10		11	
	Fast	7		11	
VA	Slow	8	3	13	4
	Average	12		10	
	Fast	10		13	
VIB	Slow	20	5	23	5
	Average	17		13	
	Fast	16		17	
VIA	Slow	6	3	4	7
	Average	5		8	
	Fast	4		3	
VIIB	Slow	34	−12	17	−1
	Average	9		5	
	Fast	−9		10	
VIIA	Slow	−23	6	−5	2
	Average	2		8	
	Fast	9		9	
VIIIB	Slow	18	25	14	17
	Average	−6		1	
	Fast	−19		−1	
VIIIA	Slow	6	20	−16	11
	Average	−11		1	
All	All	7	4	7	6

shows that the increase in absolute achievement during the second semester of 1921-22 averaged seven months in the experimental schools and four months in the control schools. Thus again it appears that the experimental schools were more efficient as regards the achievement of their pupils. As their "progress quotient" was considerably greater during this semester the gain in absolute achievement can not be attributed to holding back the pupils.

The achievement quotients at the beginning and end of the second semester of 1921-22. Table XVIII likewise presents the gains of the two groups of schools in achievement quotients. According to these quotients the gain of the experimental schools was only one point greater than that of the control schools.

The correlation of achievement and intelligence at the beginning and end of the second semester of 1921-22. The coefficients of correlation between absolute achievement and intelligence at the beginning and end of this semester were as follows:

Experimental			Control		
Beginning	End	Gain	Beginning	End	Gain
.54±.01	.75±.01	.21	.51±.01	.53±.01	.02

Judging from these coefficients it seems that there was a very decided gain in the relation of achievement to intelligence on the part of the experimental schools, but practically no change in the control schools.

The relation of the intelligence and achievement quotients at the beginning and end of the second semester of 1921-22. Table XIX, which is similar to Tables XIII and XVII, presents the relation of the achievement and intelligence quotients for the second semester of 1921-22. Comparing the figures for the beginning and end of this semester there seems to have been no noticeable change in the situation.

A comparison of the coefficients of correlation of the intelligence and achievement quotients is more favorable to the experimental schools. These coefficients were as follows:

Experimental			Control		
Beginning	End	Gain	Beginning	End	Loss
—.36±.01	—.34±.01	.02	—.39±.01	—.52±.01	.13

The change in the coefficients of the control schools showed a decrease in the adaptation of instruction to pupils of all levels of

TABLE XIX. MEDIAN ACHIEVEMENT QUOTIENTS OF THE PUPILS OF DIFFERENT LEVELS OF INTELLIGENCE AT THE BEGINNING AND END OF THE SECOND SEMESTER OF 1921–22

Intelligence Quotient	Achievement Quotients			
	Experimental		Control	
	Beginning	End	Beginning	End
150–59	98	89	85	88
140–	96	99	88	92
130–	102	106	95	101
120–	100	109	99	104
110–	102	108	104	109
100–	106	112	105	111
90–	106	115	105	112
80–	110	114	107	118
70–	111	121	113	119
60–	114	131	119	128
50–	125	125	132	131
All	105	113	104	109

intelligence, but in the experimental schools such adaptation seems to have remained about the same.

The achievements of the two groups of schools during the second semester of 1921-22 as measured by the Omnibus Test. In planning this experiment it was decided to make use of tests in reading and arithmetic because those are generally considered the two most important subjects of the elementary curriculum and further because it was believed that the results obtained would give a fairly reliable index of the achievements of the pupils in all subjects. In order to provide a partial check upon this latter assumption a test was devised by the writer and given to the pupils of grade VIB and above at the regular testing period in May, 1922. This test, which was called the Omnibus Test,[1] contained questions in geography, history, grammar, elementary science and certain phases of arithmetic not covered by the standardized tests used. The scores made on this test were translated into achievement ages and quotients in the same way as for the other tests of achievement.

Table XX presents the median ages and quotients for the various grades and sections of the two groups of schools. It may be seen from this table that the showing made upon this test by

---

[1] See Appendix C.

TABLE XX. MEDIAN GRADE AND SECTION OMNIBUS ACHIEVEMENT AGES AND QUOTIENTS OF THE PUPILS OF THE TWO GROUPS OF SCHOOLS AT THE END OF THE SECOND SEMESTER OF 1921–22

Grade	Section	Ages		Quotients	
		Experimental	Control	Experimental	Control
VIB	Slow	10–8		97	
	Average	12–3	10–10	102	88
	Fast	10–8		90	
VIA	Slow	10–8		85	
	Average	11–1	15–2	89	120
	Fast	12–2		85	
VIIB	Slow	11–10		95	
	Average	12–0	14–10	92	109
	Fast	15–3		100	
VIIA	Slow	11–1		86	
	Average	14–9	14–4	102	107
	Fast	16–8		114	
VIIIB	Slow	12–11		98	
	Average	13–5	15–7	87	108
	Fast	15–2		90	
VIIIA	Slow	11–7		88	
	Average	14–11	17–3	96	111
	All	12–8	14–11	94	107

the control schools was very much better than that made by the experimental schools. The average difference was over two years of achievement age and thirteen points of achievement quotient. Inasmuch as the pupils were not given a similar test at any previous time the relative gain can not be computed. The difference between the two groups is so great, however, that it is evidently significant. A very probable conclusion is that in the experimental schools there was a tendency to emphasize the instruction in reading and arithmetic to the neglect of that in the other subjects. This tendency was probably not due to the fact that the teachers and pupils were consciously striving to prepare to make better scores upon the tests but that merely through the use of the tests in reading and arithmetic attention was called to pupils' weaknesses in these subjects and therefore unusual care was taken to correct these weaknesses.

Individual opinion as to the relative importance of arithmetic and reading as compared with the elementary school subjects covered by the Omnibus Test will largely determine one's belief as to whether or not the experimental schools made a relative gain in achievement during the course of the experiment. Inasmuch as there was an average relative gain of only about one point per semester in the achievement quotient on the part of the experimental schools it is the opinion of the writer that there was not any greater efficiency in the achievement of this group of schools. On the other hand, he does not believe it should be asserted that in so far as achievement was concerned there was a decidedly smaller degree of efficiency.

Summary. The data presented in this chapter considering them from the standpoint of the experimental schools relative to the control schools may be briefly summarized as follows:

1. At the beginning of the experiment the median achievement age as measured by the tests used was four months lower. During the experiment slightly greater efficiency was shown, averaging about one month per semester, according to the reading and arithmetic test results. According to the results on the Omnibus Test, however, the median achievement age was twenty-seven months lower at the conclusion of the experiment.
2. The median achievement quotient derived from the arithmetic and reading tests was two points greater both at the beginning and end of the experiment. Allowing for the effect of the new entrants and eliminees, however, there was a relative gain of about one point per semester. The Omnibus Test achievement quotient was thirteen points lower.
3. The correlation of intelligence and achievement was .08 greater at the beginning and .25 greater at the close of the experiment.
4. At the beginning of the experiment instruction was somewhat better adapted to the inferior than to the superior pupils in both groups of schools. On the whole there was little change in this situation.

Considering these items together it seems that in so far as achievement was concerned the efficiency of the experimental schools was no greater than that of the control schools. The slightly greater efficiency in reading and arithmetic was at least balanced by the results of the Omnibus Test. If we assume that the measurement of achievement shows no advantage for either group of schools it may be said that the experimental plan of organization was more efficient than the traditional plan because of the fact that the progress of the pupils was considerably greater in the experimental than in the control schools. If, on the other hand, it is considered that the Omnibus Test showed a distinctly greater degree of efficiency as regards the total achievement for the control schools, this must be balanced against the greater progress made in the other group and a less definite conclusion reached.

CHAPTER V

A STUDY OF THE PUPILS WHO REMAINED IN SCHOOL THROUGHOUT THE COURSE OF THE EXPERIMENT

Although there was no reason to suspect that a study of the records of the pupils who participated in this experiment throughout the three semesters that it continued would yield results and conclusions materially different from those presented in Chapters III and IV, yet it was thought desirable to make such a study. Therefore this chapter will present certain data concerning the pupils who were in the schools in February, 1921, and remained therein until the close of the experiment. These pupils did not compose as large a group as might be expected because the pupil population of both groups of schools was very unstable. Slightly less than 60 percent of the pupils tested at the beginning of the experiment were still in the schools at its conclusion. In making a study of these pupils the tabulations were not made by separate semesters but all three semesters were taken together.

The placement of the pupils and their progress throughout the grades. Tables XXI and XXII show the percents of pupils in the two groups of schools gaining or losing the given number of semesters during the three semesters that the experiment continued. From these tables it may be seen that only 34 percent of the pupils of the experimental schools made just three semesters of progress, whereas 62 percent of those of the control schools did so. The percents making more than this amount of progress were 26 and 7, respectively, and those making less, 40 and 32. The average amount of progress made was 2.79 semesters in the experimental schools but only 2.67 semesters in the control schools. Dividing these figures by three to reduce them to a semester basis and then by the median I. Q.'s gives "progress quotients" of 93 and 89, respectively. Therefore it appears that in so far as progress was concerned the experimental plan of organization was somewhat more efficient for those pupils remaining in school throughout the experiment than was the traditional plan. The

TABLE XXI. PERCENTS OF THE PUPILS OF THE EXPERIMENTAL SCHOOLS PRESENT THROUGHOUT THE EXPERIMENT THAT GAINED OR LOST THE GIVEN NUMBER OF SEMESTERS

Feb., 1921		Semesters Gained or Lost															
Grade	Section	−5	−1	0	+1	+2	+2½	+3	+3½	+4	+4½	+5	+5½	+6	+7	+10	+11
IB	Slow				22	52	5	20		2							
	Average				11	28	1	33		25			2				
IA	Slow		1	3	9	38	1	38	8	2							
	Average				4	15		44	14	17		3			1	1	
	Fast				4			22	16	27	27	2	2				
IIB	Slow				23	51	2	15	8	2							
	Average				6	33	9	39	9	3							
	Fast					9	5	36		23	27						
IIA	Slow			1	8	59		25	4	1	1						
	Average			1	2	12		52	11	9	12						
	Fast					6		13	29	19	32						
IIIB	Slow			1	8	59		19	12			1					
	Average				6	27		47	8	2		4		4			2
	Fast				3	6	6	34	9	17	17	6	3				
IIIA	Slow			1	17	47		23	7	2	3						
	Average				7	20		39	16	10	4	3					
	Fast				2	3		32	13	22	22	2	5				
IVB	Slow				7	46		31	7	7		1					
	Average					19		36	8	25	6			6			
	Fast					2		1	1	3	2			1			
IVA	Slow			1	8	44		42	4								
	Average				2	16		41	2	20	11	5	2				
	Fast																
VB	Slow	1			5	54		26	5	5	2						
	Average					14		39	11	7	27	2					
	Fast					5		11		16	67						
VA	Slow				10	54		28	1	6							
	Average			2		22		44	13	11	8						
	Fast							20		60	13	7					
VIB	Slow				13	20		55	3	5	5						
	Average				3	14		57		12	12	2					
	Fast							13	3	7	77						
VIA	Slow			2	9	61		20	7								
	Average					20		46	18	16							
	Fast					10		10		80							
VIIB	Slow				7	60		33									
	Average					16	2	83									
	Fast																
VIIA	Slow					100											
	Average				8	92											
	Fast					100											
VIIIB	Slow																
	Average				100												
VIIIA	Slow																
	Average																
All	All	.04	.04	.4	7	32	1	34	7	10	7	1	.4	.2	.04	.04	.04

TABLE XXII. PERCENTS OF THE PUPILS OF THE CONTROL SCHOOLS PRESENT THROUGHOUT THE EXPERIMENT THAT GAINED OR LOST THE GIVEN NUMBER OF SEMESTERS

February 1921 Grade	−1	0	+1	+2	+3	+4	+5	+6
IB		1	16	51	30	3		
IA		1	7	24	55	13		
IIB			7	37	56	1		
IIA		1	4	31	60	3		
IIIB		1	6	25	64	2		3
IIIA		4	5	26	58	7		
IVB			3	16	76	3	2	
IVA			1	23	73	2		
VB			1	17	78	2		1
VA			3	17	62	17	1	
VIB				21	69	10		
VIA			2	22	64	12		
VIIB			5	20	75			
VIIA	7		7	87				
VIIIB								
All	.1	1	5	26	62	6	.2	.3

difference of four points in the "progress quotients," especially when it is remembered that this is an average difference for three semesters, is large enough to justify the above statement.

The achievements of the pupils. Table XXIII presents the median achievement ages and quotients for the two groups of schools at the beginning and end of the experiment. From these data it may be seen that the gain on the part of the pupils of the experimental schools was three months of achievement age greater than that for the other group of schools and that the gain in achievement quotient was one point greater. These figures show that for the pupils who remained throughout the course of the experiment the experimental schools were slightly more efficient in so far as achievement was concerned.

Summary. The evidence afforded by the study of the pupils who remained in school during the course of the experiment shows that for these pupils the experimental plan of procedure resulted in appreciably greater progress according to the ability of the children and in slightly greater achievement. The difference in the "progress quotients" was four points and that in the achievement quotients, one point. Thus the general conclusion to be drawn

TABLE XXIII. MEDIAN ACHIEVEMENT AGES AND QUOTIENTS OF PUPILS WHO WERE PRESENT THROUGHOUT THE EXPERIMENT IN FEBRUARY, 1921 AND MAY, 1922

1921		Achievement Age				Achievement Quotient			
Grade	Section	Experimental		Control		Experimental		Control	
		1921	1922	1921	1922	1921	1922	1921	1922
IA	Slow	6–5	7–7	6–6	9–2	101	113	91	108
	Average	5–10	9–7			88	132		
	Fast	5–6	11–6			70	125		
IIB	Slow	6–6	6–5	7–7	7–6	100	100	101	93
	Average	6–7	8–4			90	106		
	Fast	8–0	9–5			103	108		
IIA	Slow	6–7	8–11	8–1	9–6	90	112	93	105
	Average	7–6	9–8			93	111		
	Fast	8–1	9–6			91	102		
IIIB	Slow	8–2	9–5	9–1	10–8	109	118	105	104
	Average	8–10	10–0			106	107		
	Fast	8–10	8–11			98	109		
IIIA	Slow	8–11	8–8	9–8	10–0	115	105	115	98
	Average	10–1	9–5			113	99		
	Fast	9–10	10–6			107	92		
IVB	Slow	8–11	9–4	9–1	10–11	112	103	108	115
	Average	9–11	10–4			106	110		
	Fast	10–6	10–2			105	100		
IVA	Slow	9–7	10–7	10–3	12–3	107	111	102	124
	Average	10–11	11–10			111	117		
	Fast	10–6	12–10			100	111		
VB	Slow	9–7	11–8	10–7	11–11	107	116	104	114
	Average	11–0	12–8			110	115		
	Fast	10–5	12–10			98	114		
VA	Slow	10–2	12–0	11–4	12–1	106	117	107	111
	Average	11–8	13–8			108	118		
	Fast	11–11	16–10			103	137		
VIB	Slow	10–1	12–10	12–1	14–1	104	124	107	120
	Average	11–11	14–4			106	118		
	Fast	13–10	16–4			112	116		
VIA	Slow	11–7	13–2	12–1	14–5	112	111	107	112
	Average	12–6	14–0			104	114		
	Fast	15–0	14–6			112	110		
VIIB	Slow	12–5	15–2	13–11	14–4	108	123	105	108
	Average	12–4	13–7			103	117		
	Fast	14–6	15–10			103	118		
VIIA	Slow	12–6	13–7	14–7	14–9	110	110	101	113
	Average	13–7	16–1			103	114		
	Fast	15–6	18–2			102	124		
VIIIB	Slow	....	16–5	16–5	16–7	....	125	118	120
	Average	12–0	17–0			100	114		
	Fast		19–8			....	120		
VIIIA	Slow	....	14–0			....	95		
	Average	....	17–10	....	17–4	....	113	....	112
All	All	9–10	11–6	10–2	11–7	106	112	104	109

from this study is the same as that drawn from the study of all the pupils, that the experimental plan of organization was somewhat more efficient than was the traditional plan.

CHAPTER VI

A SPECIAL STUDY OF THE BRIGHTER AND DULLER PUPILS

It is evident that such an experiment as the one described in this bulletin might not have the same effect upon the efficiency of the instruction of the brighter, the average and the duller pupils. In view of this fact a special study was made of the brighter and another of the duller pupils in order to discover the effect of the experimental plan of organization upon the efficiency of the instruction of these two groups. For the purpose of the two studies the records of those pupils whose I. Q.'s as found at the first testing period were 115 or higher and of those whose I. Q.'s were less than 80 were used. The former group included about one-sixth of the total number of pupils and the latter group about one-fifth. All records not complete for the duration of the experiment were rejected so that the number of pupils actually included in these studies was reduced to 199 brighter pupils and 514 duller pupils from the experimental schools and 396 brighter and 291 duller pupils from the control schools.

The placement of the brighter pupils and their progress through the grades. Of the 199 pupils of the experimental schools 2 percent were placed in the slow sections, 23 percent in the average sections and 75 percent in the fast sections at the beginning of the experiment. When it closed the respective percents were 1, 51 and 49. The marked reduction of the number in the fast sections was due to the fact that by the close of the experiment these pupils had gained one semester or more and in many cases were not quite bright enough to attempt to make further gain, at least immediately.

Table XXIV shows that the number of semesters gained by the brighter pupils of the experimental schools varied from two to six, and by those of the control schools from one to four. The percents of the brighter pupils of the experimental schools making less than regular, regular and more than regular progress, were 9, 23, and 68, respectively. In the control schools the corresponding

TABLE XXIV. PERCENTS OF THE BRIGHTER AND OF THE DULLER PUPILS OF THE TWO GROUPS OF SCHOOLS GAINING OR LOSING THE GIVEN NUMBER OF SEMESTERS DURING THE EXPERIMENT.

	−5	0	1	2	2½	3	3½	4	4½	5	5½	6	7	10	11	Average Progress	"Progress Quotient"
Brighter Pupils																	
Experimental																	
Slow.......						25	25	50								3.63	97
Average....				13		42	22	11	7	2	2					3.29	92
Fast.......				5	2	17	9	19	40	3	3	2				3.97	106
All........				7	2	23	13	18	32	3	2	2				3.80	102
Control			2	15		72		11								2.92	77
Duller Pupils																	
Experimental																	
Slow.......	.2	2	14	57	.2	20	4	3	.5	.2						2.14	99
Average....			13	27	2	43	3	8	13			1*	1*	1*	1*	2.91	133
Fast.......			13	25			13	38	1							3.13	149
All........	.2	1	14	54	.4	23	4	4		.2		.2*	.2*	.2*	.2*	2.26	105
Control......		1	8	36		50		5		.3		1*				2.56	119

*These large amounts of progress were made by foreign-born pupils who, at the beginning of the experiment, were so handicapped by their inability to use the English language that they made low test scores and did poor school work. Many of these pupils were able to skip the work of several semesters as soon as the language difficulty was overcome.

figures were 17, 72 and 11. Table XXIV also shows that the average progress of the brighter pupils of the experimental schools was .88 of a semester greater than that of the pupils of the control schools and that their "progress quotient" was twenty-five points greater. These differences show that the experimental schools were much more effective in so far as the rate of progress of the brighter pupils was concerned.

The achievements of the brighter pupils. The table just below gives the median achievement ages and quotients of the brighter pupils of both groups of schools in February, 1921, and May, 1922.

	Experimental			Control		
	1921	1922	Gain	1921	1922	Gain
Achievement Age	11–4	12–11	19	11–2	12–7	17
Achievement Quotient	103	112	9	98	104	6

From these data it is evident that the more rapid progress of the pupils of the experimental schools did not result in a lessening of their relative achievement but was accompanied by a small gain. This gain in relative achievement amounted to two months in terms of achievement age or three points in terms of achievement quotient. Thus considering progress and achievement together, it may be said that for the brighter pupils the experimental plan of organization resulted in a marked increase of efficiency.

The placement of the duller pupils and their progress through the grades. Of the 514 duller pupils from the experimental schools, 86 percent were placed in the slow sections, 12 percent in the average and 2 percent in the fast sections at the beginning of the experiment. At the close of the experiment the respective percents were 85, 13 and 2. Thus it is apparent that there was practically no change in the number of pupils in each of the three sectional groups.

Table XXIV shows that the number of the duller pupils making more than normal progress was not very large in either group of schools. Slightly over one-half of the duller pupils of the experimental schools made regular progress in the slow sections, which resulted in their covering two semesters' work during the three semesters of the experiment. Slightly less than one-fourth of them made three semesters' progress by maintaining membership in the average sections. Ten percent managed to make more than normal progress, while 15 percent made less than two semesters. In the control schools 50 percent made normal progress, 45 percent less and 6 percent more. The average progress was three-tenths of a semester greater for the pupils of the control schools and the "progress quotient" fourteen points greater.

Analyzing the data presented above it is apparent that more of the duller pupils were failed in the control schools than in the experimental schools. Since, however, pupils were able to advance in the slow sections without failure while covering less than the normal amount of work the average progress was less in the experimental schools. As was true in the case of the brighter pupils more of the pupils from the experimental schools made extra progress.

The achievements of the duller pupils. The table just below gives the median achievement ages and quotients of the duller pupils of both groups of schools in February, 1921, and May, 1922.

	Experimental			Control		
	1921	1922	Gain	1921	1922	Gain
Achievement Age	8–10	10–6	20	9–3	10–9	18
Achievement Quotient	112	114	2	115	114	–1

These data show that the gain made by the duller pupils of the experimental schools was two months of achievement age greater than that made by those of the control schools and that their gain in achievement quotient was three points greater. Therefore it can be said that in so far as achievement was concerned the experimental plan of procedure was slightly more efficient for the duller pupils than the traditional plan used in the control schools.

Summary. A special study of the brighter and duller pupils who were in school throughout the experiment yields the following results and conclusions:

1. The brighter pupils of the experimental schools had a "progress quotient" twenty-five points greater than did those of the control schools.
2. The relative gain of the brighter pupils of the experimental schools in median achievement quotient was three points.
3. The "progress quotient" of the duller pupils of the experimental schools was fourteen points less than that of the duller pupils of the control schools.
4. The duller pupils of the experimental schools made a relative gain of three points in their median achievement quotient.

A fair statement of the conclusions to be drawn would seem to be that the experimental plan of organization was considerably more efficient than the traditional plan in so far as it concerned the brighter pupils, but that in the case of the duller pupils it was somewhat less efficient.

CHAPTER VII

RESULTS AND CONCLUSIONS

A brief statement of the results of this experiment. The results actually obtained in this experiment may be listed as follows:

I. At the beginning of the experiment the placement of the pupils involved a relative loss in placement of .05 semester on the part of the experimental schools. This and the other amounts of progress are computed relative to the capacity of the pupils.

II. The main study, which included all the pupils of the two groups of schools, showed that:

1. The average progress was .06 semester larger in the experimental schools than in the control schools.
2. There was a relative gain for the experimental schools of about one point per semester in the achievement quotient as measured by the arithmetic and reading tests.
3. The achievement quotient derived from the Omnibus Test was thirteen points less for the experimental schools.

III. A special study of the pupils who remained in school throughout the experiment gave the following results:

1. The average progress for the experimental schools was .04 semester greater than that for the other group.
2. There was a relative gain for the experimental schools of one point in the achievement quotient.

IV. A special study of the brighter pupils revealed the following facts:

1. Those of the experimental schools progressed at a rate .25 semester greater than did those of the control schools.
2. The relative gain in the achievement quotient on the part of the pupils of the experimental schools was three points.

V. A special study of the duller pupils gave the following results:

1. Those of the experimental schools made, on the average, .14 semester less progress per semester than did those of the other group of schools.
2. There was a relative gain of three points in the median achievement quotient for the experimental schools.

It seems fair to summarize these results by saying that for pupils of all degrees of intelligence combined the experimental plan of organization was more efficient as regards progress and about the same as regards achievement, as compared with the traditional plan. The difference in progress was considerably more than enough to balance the relative loss caused by the placement of the pupils at the beginning of the experiment.

Conclusions to be drawn from these results and their application to school systems in general. The comparisons that were made between the schools taking part in this experiment and certain other city school systems seem to show that the results obtained in this experiment and the conclusions based thereon are fairly applicable to school systems in general. Assuming that this conclusion is warranted, the question remains as to whether or not the classification of pupils along lines similar to those followed in this experiment should be recommended to school administrators as a practical method of procedure. In considering this question it should be recognized that the public school superintendent or supervisor can ordinarily exercise a somewhat higher degree of supervision over the schools under his control than could the writer over the schools participating in this experiment. Therefore, it should be possible to secure somewhat more favorable conditions for carrying out the experimental plan of organization than were possible in this experiment.

In the second place, the question arises as to how large a gain in output, that is to say in progress and achievement, is required to justify a certain amount of additional investment. In this experiment the cost in both money and time was considerably larger per pupil than would be necessary in the usual public school situation. Ordinarily pupils would not need to be tested so often nor would it be necessary to use tests of achievement. Furthermore, there were many tabulations and computations made in this project that would not be necessary in the ordinary school situation. The cost of group intelligence tests is only a few cents per pupil, in some cases being as low as one and one-half cents and in few more than ten cents. If the teachers scored the papers there would be no extra expense involved therein. Thus the cost of the tests and a rather small amount of clerk hire would be all the unusual outlay required to make use of group intelligence tests for purposes of placing pupils. Certain plans of doing this have involved a decrease in the average number of pupils per teacher or per room or some other element of additional investment. In this experiment there was no such expenditure, nor need there be in the usual situation. The desirability of reducing class size, whether in this or some other type of organization, is a separate problem. Therefore the total cost of the type of organization used in the experimental schools amounts to only a fraction of one percent of the total expenditure per pupil. As the gain in progress on the part of the experimental schools amounted to several percent of the total progress and as there was no loss in achievement, and, furthermore, as it is probable that under ordinary conditions the gain would be greater than it was in this project it would seem that an additional investment of a fraction of one percent would be entirely justifiable.

There remains, however, another point that must be considered in this connection. In Chapter I, output was defined as being composed of progress and achievement. There are undoubtedly other less tangible factors that constitute a part, and a rather important part, of the output of a school system. Such outcomes as industry, good citizenship, intellectual honesty, social development, etc., were either not measured in this experiment or measured so indirectly that no assumptions can be made concerning their presence and amount. This fact does not invalidate the conclusions reached, but merely signifies that these other outcomes of instruction must be considered in their interpretation. The fact that we cannot measure the total output should not bar us from measuring that which can be measured nor from proceeding according to what our measurements reveal until more complete measurements are possible.

It must also be remembered, as was stated in Chapter I, that there were really two problems involved in this experiment. It is possible that a portion or all of the results obtained in this experiment might be secured in a somewhat similar experiment in which the pupils were classified according to teachers' judgments. Especially might this occur if the teachers participating were well-trained and experienced, and perhaps had given special study to the problem of classifying pupils according to their capacities. There were several reasons why the pupils in the control schools were not so classified, the chief one being that it was impracticable in the given situation. It may be suggested that since fast, average and slow sections were not formed in the control schools, they should not have been formed in the experimental schools. As was stated in Chapter I, the use of intelligence tests for the purpose of placing pupils implies that the pupils be placed according to their capacities and that it would not have been possible to arrange an experiment that would show the value of intelligence tests for the purpose mentioned unless such sections had been formed. Also the writer does not believe that the classification of the pupils of the experimental schools according to the teachers' judgments would have yielded as favorable results as did their classification according to the principles enumerated in Chapter I. This belief is based upon a study of the accounts of various experiments and of the teachers' estimates of capacity and the average school marks actually given in this experiment. These disagreed with the results of the intelligence tests in many cases and in most of these the latter appeared to furnish a more reliable means of predicting future progress and achievement than did the former.

In considering the conclusions reached from this study it should be borne in mind that the total time included was only three semesters. It is probable that if the experiment had continued for a longer time, say for eight or ten years, certain effects would have been noted that did not appear during the three semesters or effects that were present might have appeared in much more pronounced fashion. In general it seemed that as the experiment progressed from semester to semester the plan of organization being tried out gave better results. If the teachers had had several years' experience with such a plan the results might have been still more favorable. The plan was new to the teachers and hence they probably could not do their best work at first. On the other hand, it is possible that a division of the pupils into three groups might tend to make the teachers feel less responsible for the achievements of the pupils, especially those of the duller ones. They might more or less unconsciously come to feel that the pupils placed in the slow sections could not be expected to do a very high quality of work and that therefore they were not worth much attention and effort. Such a result would, of course, be decidedly undesirable.

Considering the facts and possibilities mentioned above it is the opinion of the writer that the use of intelligence tests as the chief basis of classifying pupils increases the output of the school sufficiently to justify the additional expense involved. It is not, however, a panacea for all inefficient schools nor a method of organization that should be rushed into by every school administrator before he has made a careful study of its installation and operation.

APPENDIX A

A COMPARISON OF THE PUPILS ENTERING AND LEAVING SCHOOL DURING THE EXPERIMENT WITH THE TOTAL NUMBER OF PUPILS

Necessity for this comparison. Inasmuch as the shifting of membership within both groups of schools was so large, it seemed wise to take definite account of its effect upon the results and conclusions reached. In Chapters III and IV, where these results and conclusions are given, this effect has been considered. It was more or less probable that the number or mental capacities of the pupils eliminated from the experimental schools might be considerably influenced by the conditions of the experiment. For example, the recognition of the ability of the brighter pupils might tend to hold a larger percent of them in school and the placing of the duller pupils in slow sections might cause more of them to leave school than would normally be the case. This would, of course, materially raise the general mental level of the pupil material. On the other hand, it is possible that by placing many duller pupils, who would otherwise be failed, in the slow sections more of them would be held in school and that by allowing the brighter pupils to progress more rapidly they would be encouraged to leave school sooner than would otherwise be the case. Such results as these would lower the general mental level. Or perhaps some other combination of the four possible results just mentioned took place, so that more pupils of all degrees of ability were held in school, or more eliminated. Or again, other effects than those mentioned might have resulted. In regard to the new entrants, a priori reasoning would lead to the conclusion that they would have no effect upon the outcome of the experiment, since its operation would not in any way cause them to enter or not to enter school. However, it was thought best to make a study of them as well as one of the eliminees.

The effect of the pupils entering and leaving school during the experiment upon the total school population. It was found that during each of the three semesters of the experiment the percent of pupils eliminated from the experimental schools was much greater than that from the control schools, the averages being about 12 and 7 percent, respectively. It might seem, therefore, that the experimental plan of organization resulted in increasing the amount of elimination. The writer does not believe, however, that this was the case. If it had been, the elimination rate for the pupils in the different sectional groups probably would have varied considerably. A study of this phase of the question shows that for each of the semesters the percents of all the pupils belonging to the fast, average and slow sections that were eliminated were practically the same. To word it differently, the percent of all pupils eliminated that had been in the fast sections was almost exactly the same as the percent of all pupils placed therein. A similar condition held for the other sections. Furthermore, the principals of the experimental schools stated that the elimination was no greater than was usual.

Table XXV shows the effects of the entrance and the elimination of pupils upon the total school population. It is to be read as follows, taking the first double column of the row of entries following "Med. Chron. Age" as an example: the elimination of pupils during the second semester of 1920-21 caused a decrease of one-tenth of a year more in the median chronological age of the pupils of the experimental schools than in that of the control schools. The entrance of new pupils during this time had no effect.

TABLE XXV. THE EFFECTS OF THE ENTRANCE AND ELIMINATION OF PUPILS UPON THE TOTAL PUPIL POPULATION

	Second Semester of 1920–21		Summer of 1921		First Semester of 1921-22		Second Semester of 1921-22	
	Elim.	NewE.	Elim.	NewE.	Elim.	NewE.	Elim.	NewE.
Med. Chron. Age........	−.1		−.2		−.1	−.1	−.2	
Percent Accelerated......	−1.		−1.	+1.			−1.	
Percent Retarded........	+1.		+1.				+1.	
Aver. Retardation.......	−.03	+.03	−.04		−.02	+.02	−.04	
Aver. Progress..........		−.01		−.01		−.01		
Prog. Quotient..........		−1.						
Median M. A............			−.1			−.2	−.1	
Median I. Q............						−1.	+1.	
Median A. A............			−2.	−1.		−1.		
Median A. Q.				−1.				

In making use of the data in this table it must be borne in mind that all of the eliminated pupils were not included in the tabulations from which the data were derived. In a rather large number of cases the individual record cards of pupils who had left school were not returned to the writer along with the cards of those still in school. Practically all of these cases were in the control schools. In other cases the pupils were absent at the time of testing but did not actually withdraw from school until later, not returning to be tested in the meantime, so that another possible source of discrepancy was introduced. In view of these facts it was not certain that the effects listed in the table were all of the effects or were the true effects produced upon the pupil material by the pupils who left during the experiment. In the case of the new entrants there were no such opportunities for records to be lost unless the pupils concerned not only entered but left during the same semester, in which case they would not have been included in the tabulation.

The effect of the differences between the new entrants and eliminees and the total pupil population in so far as they relate to progress were considered in Chapter III. On the whole these effects were comparatively small. Those having to do with achievement were not used in Chapter IV or elsewhere. The reason for this was that all the tabulations in that chapter were made for the pupils who were present throughout the semester and hence did not need to be included for the pupils entering or leaving during the given semester. They are merely presented here as a matter of interest.

APPENDIX B

THE RELIABILITY AND CORRELATION OF THE TESTS USED IN THIS EXPERIMENT

In considering the results of such an experiment as the one described in the body of this report the question of the reliability of the tests used at once arises. The writer will not go into the matter in a detailed way but will merely present such coefficients of correlation and other measures of the reliability of the tests as were obtained and comment briefly thereon. No attempt was made to compute all the possible correlations between the tests used.

Constant and variable errors. Before proceeding to give the data referred to in the preceding paragraph, a brief discussion of the errors present in test scores seems appropriate. These errors may be classified as constant and variable.

Constant errors are those which are the same or approximately so for the group being tested. If, for example, the person giving the test allows less time than the directions call for a constant error is introduced, the effect of which is to lower the scores of all pupils taking the test. On the other hand, if too much time is allowed the scores are too large. Probably the most frequent constant errors are those due to what is often called "practice effect." If a duplicate form of a test is given the scores made thereon are ordinarily somewhat higher than those made at the first trial. Such constant errors were, of course, present in this experiment but as they were equally present for the two groups of schools it was not necessary to make any allowance for them.

Variable errors are those which differ for the different individuals taking the test. They are due to a number of causes. On any given day certain pupils are below par physically or mentally and therefore are likely to make a lower score than they would ordinarily. Such happenings as the breaking of a pencil point, the dropping of a test paper upon the floor or some occurrence distracting an individual's attention cause variable errors. All of these mentioned so far result in lower scores. On the other hand, it may be that the particular form of a test used contains items which happen to be well known by a few members of the group taking the test. Such a condition results in an increased score. Scores may also be increased if a pupil turns the page and starts before the signal is given, if he does not know the correct answer but gets it by looking at someone else's paper, and by various other causes. It is usually impossible to determine the variable errors present in the scores of the individual pupils, although this can sometimes be done by a more or less detailed investigation. The effect of these errors is that the scores of many of the pupils are slightly too large or too small and those of a few are very much in error. On the other hand, the variable errors cause very little or no change in the average. In the long run they are as often positive as negative and therefore offset each other in the computation of averages.

The reliability of the Pressey Primer and the Illinois General Intelligence Scales. As the two scales named were the only ones used more than once in this experiment, they are the only ones for which the reliability can be calculated. The coefficients of correlation or of reliability,[1] the indices of reliability,[2] the probable errors of measurement,[3] and the percents these probable errors were of the respective medians were calculated.[4] Throughout the discussion of these measures of reliability it should be remembered that they were all computed from the use of tests at intervals of about six months and one year and therefore should not be expected to show as high a degree of reliability as if the time intervals had been shorter. In most studies of the reliability of tests the interval between the periods at which the tests were given has not exceeded a few days.

---

[1] The coefficient of correlation between repetitions or duplicate forms of the same test is called the coefficient of reliability.

[2] The index of reliability is the square root of the coefficient of reliability. It measures the correlation between the score on one trial of a test and the true score. This true score is the average of the scores made upon an infinite number of trials of the test after these have been corrected for any constant errors.

[3] The term "probable error of measurement" bears the same relation to the index of reliability that the probable error of estimate bears to the coefficient of reliability. It is a measure of the variable error by which a pupil's score upon one trial of a test deviates from his true score. The formula is $.6745\sigma\sqrt{1-r}$. For σ the average of the standard deviations obtained from the scores made on each of two trials is used.

[4] The complete tables are to be found in the dissertation by the same title and author.

Table XXVI, Part A, shows that there was in general little difference in degree of reliability between the Pressey Primer and the Illinois General Intelligence Scale, that of the former being slightly higher. The average coefficient of reliability was in each case about four-tenths for the single half-grade groups and not far from seven-tenths for all grades combined. The average indices of reliability were somewhat greater than six-tenths and eight-tenths, respectively. The probable error of measurement averaged about nine points, or 15 percent of the median, in both cases. In the case of the Illinois Scale this amounts to almost one year of mental age, whereas in that of the Pressey it is somewhat less.

TABLE XXVI. DATA CONCERNING THE INTELLIGENCE TESTS USED IN THIS PROJECT

A. Reliability

	Coefficient of Reliability	Index of Reliability	Probable Error of Measurement	$\frac{\text{P.E. Meas.}}{\text{Median}}$
Pressey (Nov., 1920 and May, 1921)				
Grade Average	.46±.02	.67±.02	8	.16
Grades Combined	.65±.01	.81±.01	9	.15
Illinois (Form 1 in Nov., 1920 and Form 2 in May, 1921)				
Grade Average	.38±.03	.62±.03	9	.15
Grades Combined	.69±.01	.83±.01	9	.15
(Form 2 in May, 1921 and Form 1 in Dec., 1921)				
Grade Average	.46±.04	.67±.02	8	.14
Grades Combined	.73±.01	.85±.01	9	.13
(Form 1 in Nov., 1920 and in Dec., 1921)				
Grade Average	.32±.04	.55±.03	9	.14
Grades Combined	.62±.01	.79±.01	10	.15

B. Correlations Between the Different Tests Used.

	Grade Average	Grades Combined
Pressey and Dearborn (Used at Same Time)	.59±.02	.78±.01
Illinois and National (" " " ")	.63±.02	.81±.01
Pressey and Myers (" one semester apart)	.29±.03	.39±.01
Kingsbury and Myers (" " " ")	.47±.01	*
Pressey and Illinois (" " " ")	.20±.03	*
Dearborn and Illinois (" " " ")	.23±.03	*
Pressey-Dearborn and Illinois (Used one semester apart)	.36±.03	*
Dearborn and Myers (Used one year apart)	.38±.04	.52±.01
Pressey and Myers (" " " ")	.32±.04	.46±.01

*In these cases the correlations from only one grade are available.

In other words, the mental ages derived from a single application of the tests would be within that distance of the true mental ages in only about 50 percent of the cases.

Certain data as to the reliability of these two scales have been given by their authors. The administration of the Pressey scale to 365 first, second and third grade pupils gave an average coefficient of reliability of .92 between the first and second halves of the scale.[5] With two other groups of pupils numbering slightly over 100 each, coefficients of .89 and .92 were obtained.[6] The probable error of measurement was found to be between two and three points on the scale. These coefficients are naturally much higher and the probable errors much less than those obtained in this experiment because of the difference in the intervals between testing. The coefficients of reliability for Forms 1 and 2 of the Illinois scale are not quite as high as those between the two halves of the Pressey scale. Results based upon about 1000 children gave an average coefficient of .83 for grades III to VIII and one of .92 for the grades combined.[7] The probable error of measurement was between five and six points on the scale. These figures also show a considerably higher degree of reliability than do those obtained by testing at intervals of six months and one year. Inasmuch as the scale of the Illinois is finer than that of the Pressey, the probable errors are not far from the same when converted into mental ages.

The coefficients of reliability that are given for two or three other group intelligence tests run from about .75 up.[8] They tend

---

[5] Pressey, L. W. "A Group Scale of Intelligence for Use in the First Three Grades." Journal of Educational Psychology, 10, 297-308, September, 1919.

[6] Pressey, L. W. "A Group Scale of Intelligence for Use in the First Three Grades." Journal of Educational Research, 1, 285-94, April, 1920.

[7] Monroe, W. S. "The Illinois Examination." University of Illinois Bulletin, Vol. 19, No. 9, Bureau of Educational Research Bulletin No. 6. Urbana: University of Illinois, 1921. p. 47-49.

Monroe, W. S. and Buckingham, B. R. "The Illinois Examination I and II. Teacher's Handbook." Bloomington: Public School Publishing Company, 1920, p. 31.

[8] Colvin, S. S. "Educational Tests at Brown University." School and Society, 10, 27, July 5, 1919.

Colvin, S. S. "Some Recent Results Obtained from the Otis Group Intelligence Scale." Journal of Educational Research, 3, 1-12, January, 1921.

Otis, A. S. "An Absolute Point Scale for the Group Measurement of Intelto average about .80. Therefore, if these few are typical of similar tests in general, it would seem that the Pressey and Illinois scales are more reliable than are most group intelligence tests. From such a comparative standpoint coefficients of reliability around .90 and probable errors of measurement of two and five points may be said to be rather satisfactory. The differences between these figures and those obtained in this project may be largely, if not entirely, attributed to the difference in the time elapsing between the giving of the tests.

The correlations between the different group intelligence tests used in this experiment. Part B of Table XXVI presents the correlations obtained between the different tests used. It will be seen that the correlation between the Pressey scale and the Dearborn tests and that between the Illinois scale and the National tests are fairly high. An average correlation of about .60 when pupils are taken by half-grade groups and of about .80 for all grades combined is higher than is usually found between group intelligence tests.

The correlations between the results of the tests used at intervals of six months and one year are considerably lower. This would, of course, be expected as they take account not only of the differences between the tests but also of changes in the true mental abilities of the pupils during the period elapsing between the giving of the tests and of differences in the general conditions of testing at the two times. On the whole, these correlations do not compare unfavorably with similar correlations obtained elsewhere.

The writer collected data concerning the correlations found between different intelligence tests in some fifty cases. In practically all of these the different tests were given within a comparativly short time of each other, usually within the same week. The unweighted average of the coefficients of correlation was .62, which is only slightly higher than the average correlation by half-grade groups given in Part B of Table XXVI and much lower than that

---

ligence." Journal of Educational Psychology, 9, 333-47, and 237-61 May, 1918, and June, 1918.

Snarr, O. W. "Reliability of General Intelligence Tests in Classifying High School Pupils." Unpublished Thesis, University of Chicago, Chicago, June, 1919.

obtained for the grades combined. This is true although in a number of cases the coefficients were based upon several grades combined. Only about a dozen of the fifty are as high or higher than those of .78 and .81 which were obtained in this experiment when the grades were combined. In only one case was there a coefficient found higher than .90. Thus it may be said that the correlation between the Pressey scale and the Dearborn tests and that between the Illinois scale and the National tests were rather satisfactory as compared with similar correlations obtained in other experiments.

Although the coefficients given in Part B of Table XXVI were obtained from testing at intervals of one and two semesters, yet some of them compare favorably with a number of those given in the accounts of other experiments. When several half-grade groups were combined the coefficients averaged about .46.

The degree of reliability of single test scores was of concern in placing the individual pupils, but in measuring the results of the experiment this was not a matter of importance. The average used in most cases was the median, and for this the probable error is 1.25 (approx.) times the probable error of the distribution divided by the square root of the number of cases.[9] As the number of pupils included in this experiment was so large, the distribution would have had to be very scattering and the probable errors very large to cause the medians to be unreliable to any considerable degree. The distribution of the 3615 November, 1920, scores upon the Illinois scale, for example, had a probable error of 41 points, or 4.1 years of mental age. The probable error of the median was therefore about .85 point or one month.

---

[9] Yule, G. U. "An Introduction to the Theory of Statistics," London: Charles Griffin and Company, 1919, p. 338.

APPENDIX C

THE OMNIBUS TEST

As was mentioned in Chapter IV, a test called the Omnibus Test was devised by the writer to measure certain achievements of the pupils in the upper grades that were not covered by the reading and arithmetic tests used. This test was of the true-false type. It included seventy-five statements of which approximately half were correct and half incorrect. The following gives the first ten statements of the test:

1. Russia produces a large amount of wheat.......... | ....
2. The ancient Greeks were famous for their art.... | ....
3. Charcoal is made from wood.......... | ....
4. 4.6 is 100 times .46.......... | ....
5. A paragraph should be indented.......... | ....
6. Italy raises a great deal of flax.......... | ....
7. The Roman Empire was not as powerful as Greece.......... | ....
8. Digestion begins in the mouth.......... | ....
9. 41/1000 = .41.......... | ....
10. A compound sentence has at least two independent clauses.......... | ....

Every fifth statement had to do with the same subject, the five subjects included being geography, history, elementary science, arithmetic and grammar. The fifteen statements dealing with each subject were divided approximately equally between the six semesters of work covered and were in all cases based upon material mentioned in the outline of the Chicago course of study. The seventy-five statements were preceded by explicit directions and preliminary practise statements. The pupils were instructed to place a plus mark after those statements that were correct and a minus sign after those that were incorrect. A time limit of four minutes was placed upon the test.

BULLETIN NO. 13

BUREAU OF EDUCATIONAL RESEARCH
COLLEGE OF EDUCATION

THE STATUS OF THE SOCIAL SCIENCES IN THE HIGH SCHOOLS OF THE NORTH CENTRAL ASSOCIATION

by

Walter S. Monroe
Director, Bureau of Educational Research

and

I. O. Foster
Instructor, University High School

PRICE 50 CENTS

PUBLISHED BY THE UNIVERSITY OF ILLINOIS, URBANA
1922

UNIVERSITY
OF ILLINOIS
PRESS

TABLE OF CONTENTS

Preface........ 5
Chapter I. Introduction........ 7
Chapter II. Summary of Questions........ 11
Chapter III. Summary by Subjects........ 26
- *(a)* Ancient History........ 26
- *(b)* Ancient and Medieval History........ 27
- *(c)* Medieval History........ 28
- *(d)* Medieval and Modern History........ 29
- *(e)* Modern History........ 30
- *(f)* English History........ 31
- *(g)* General History........ 31
- *(h)* American History........ 32
- *(i)* Community Civics........ 33
- *(j)* Civics........ 33
- *(k)* Vocational Civics........ 34
- *(l)* Economics........ 35
- *(m)* Sociology........ 36
- *(n)* Social Science........ 37
- *(o)* Educational Guidance........ 37
- *(p)* Vocational Guidance........ 38

PREFACE

This report presents a summary of the information secured by means of a questionnaire addressed to the teachers of Social Science of the high schools of the North Central Association of Colleges and Secondary Schools. The authors have confined the report to a mere summary of the facts. Certain tabulations of the replies have not been included because it was felt that they were not sufficiently important to justify such detailed presentation. No effort has been made to interpret the facts or to summarize the comments made by a number of the persons who answered the questions.

Credit should be given Mr. Foster for originating the study and for preparing the questionnaire blank. He has also had a large part in the preparation of the manuscript for publication. The Bureau of Educational Research desires to acknowledge its indebtedness to all who cooperated by furnishing the information asked for.

WALTER S. MONROE, *Director.*

December 1, 1922.

THE STATUS OF THE SOCIAL SCIENCES IN THE HIGH SCHOOLS OF THE NORTH CENTRAL ASSOCIATION

CHAPTER I.

INTRODUCTION.

The purpose and method of this investigation. This study was undertaken for the purpose of collecting facts pertaining to the place occupied by the various social sciences in the curriculums of the high schools of the North Central Association of Colleges and Secondary Schools. In addition, certain questions relating to the general procedure of instruction were included. The specific questions asked are given on p. 8. The questionnaire blank was sent to 1273 secondary schools belonging to the North Central Association, and approximately one-half of these blanks were returned. Some were improperly filled out and had to be rejected. Others arrived too late to be included in the tabulations. The tables on the following pages are based upon the total of 475 replies.

The questionnaire blank. The questionnaire blank was printed as a four-page folder—eight and one-half by eleven. A portion of it, the general explanation, and the specific directions for answering each of the questions are reproduced here. Additional columns were added so that the following sixteen social science subjects were included in the blank:

Ancient History
Ancient and Medieval History
Medieval History
Medieval and Modern History
Modern History
English History
General History
American History
Community Civics
Civics
Vocational Civics
Economics
Sociology
Social Science
Educational Guidance
Vocational Guidance

Name...Date.................................

School...Address.......................................

	Ancient History	Ancient and Medieval History	Medieval History
1. In what years and semesters are these subjects offered?			
2. If pupils from other years are admitted, state the years			
3. How many pupils were enrolled last semester?			
4. How many pupils are now enrolled?			
5. Is this a required subject?			
6. Is this subject generally liked by the students?			
7. Does the class study the entire period covered by the text?			
8. If not, what are the date limits?			
9. Is notebook work required?			
10. Is the stereopticon used?			
11. Do you plan to offer this course again?			
12. Check the courses in which current events are taught			
13. How many weeks are required to complete this course?			
14. How many minutes per week are spent in recitations?			
15. Is supervised study used often?			
16. Is the socialized recitation commonly used?			
17. How long have these subjects been offered?			
18. Check any courses that have proved unsuccessful			
19. Do you feel that more social science should be taught?			
20. Check the courses that you feel the pupils should be required to take?			
21. Give on line 1 for each subject: the author, the basic text, the date of its copyright, and the publisher. If supplementary texts are used give on line 2 the same information for the *most important.*	2 1	2 1	2 1

Explanation and direction for answering questions. On the inside of this folder a number of questions are asked concerning social science studies. Ancient History, Ancient and Medieval History, Medieval History, Medieval and Modern History, Modern History, English

History, Civics, Economics, and Sociology need no explanation. General History is a course from the beginning of history to the present time. American History may be only the history of the United States or it may be that of the entire western hemisphere. Community Civics is a course of civics that puts most of its emphasis on the local community. Vocational Civics acquaints the pupils with various callings or vocations. A detailed study of each may or may not be made. Social Science is a mixture of any of the social studies. It is to the social sciences what General Science is to the natural sciences. Educational Guidance here means a specific course which is intended to direct the pupil's studies along lines most profitable to him; it opens up and explains new fields to him. Vocational Guidance is the direction of the pupil toward those vocations which best suit his abilities and interests. The blanks are left for you to designate any other social study which your school offers.

Each question is explained below and specific directions are given for answering it. The numbers refer to the questions. It is suggested that all the blanks for question 1 should be filled in before you pass to question 2, etc. Each question should be answered under each subject that is offered by your school unless the following directions state otherwise.

1. Put the year in Roman numerals and the semester of the year in Arabic numerals. If Civics is offered only to Seniors in the second semester, under Civics and opposite the first question you will write IV-2, etc.

2. If pupils from other years or semesters are not permitted to take any subject, write under it and opposite question 2 the word "No." If all are admitted, write "All." If the course is open only to certain years or semesters other than expressed in 1, explain by giving the years and semesters, as in question 1.

3. Give the numbers as nearly exact as possible, as 348.

4. Give the numbers as nearly exact as possible, as 125.

5. Answer in the appropriate space, "Yes" or "No."

6. Answer in the appropriate space, "Yes" or "No."

7. Answer in the appropriate space, "Yes" or "No."

8. This applies only to history. If in Ancient History the work ends at 476, write "to 476" in the appropriate space. If the work in Medieval History extends from 800 to 1700, write in the appropriate space "800-1700." It is very important that this question be answered. English History, American History, and General History are assumed to cover the traditional period; if not, state your practice.

9. Answer "Yes" or "No."

10. Answer "Yes" or "No."

11. Answer "Yes" or "No."

12. Put a check (✓) or a cross (×) in the appropriate spaces.

13. Give the exact number, as 36.

14. Give the exact number, as 250.

15. Answer "Yes" or "No."

16. Answer "Yes" or "No."

17. Answer in numerals, as "2½" for two and one-half years. If the subject has been offered longer than three years, answer "3."

18. Either a check or a cross may be used. If the cause of failure is known, please state it on the last page of this report.

19. Answer "Yes" or "No."

20. Check, as in 12 and 18.

21. This question is very important. If in Ancient History the basic text is "Breasted, Ancient Times," write "Breasted, Ancient Times, 1916, Ginn and Company." If Webster's Ancient History is used for supplementary work, write on line 2 "Webster, Ancient History, 1913, D. C. Heath." Count as supplementary texts only books covering the whole subject, which are not library reference books.

Any additional information or remarks may be placed on the last page of this sheet. Any experience that will throw light on this problem is welcomed. Any syllabus or course of study that you may send us will be much appreciated.

General plan of tabulating the replies. As the replies were received from high schools varying widely in total enrollment, the following groups based upon total enrollment were formed:

Group	Enrollment	Number of schools
I	1–150	122
II	151–400	197
III	401–800	83
IV	801–above	73
		Total........475

When no entry was made in the column for a given subject it was assumed that this subject was not taught in the high school reporting. The number of high schools in each group not teaching a given subject is found in the last column of the first five tables. Although these numbers are not repeated in subsequent tables it is necessary to keep in mind the fact that the different subjects are not taught in all of the high schools. In a number of instances some of the questions were answered for a given subject but no entry made for other questions. In such cases it was assumed that the subject was taught in the high school but for some reason the one replying to the questionnaire blank either was unable to answer the particular questions or had overlooked them. In the first five tables column nine gives the number of schools not answering the first question for each subject. Similar columns are not given for all of the tables. In the case of those tables which consist of percents apparent discrepancies due to the failure of entries to total one hundred are caused by the fact that the question was not answered by certain schools reporting. In order to economize space the percent of such schools was omitted. However, in any case it can be obtained by adding the percents given and subtracting their sum from one hundred.

General plan of report. It is interesting and illuminating to consider the information collected from two points of view: first, by questions, and second, by subjects. For obvious reasons it was necessary to make the tabulations by questions. The more important tables are reproduced on the following pages. These will be described briefly in Chapter II and certain significant facts for each subject pointed out in Chapter III.

CHAPTER II.

SUMMARY OF QUESTIONS.

Status of social sciences in secondary schools. The answers to the first four questions[1] indicate the place occupied by the various social sciences in the curriculums of the secondary schools of the North Central Association. Tables I, II, III, IV, and V give the semester or semesters in which each of the subjects is taught. For example, thirty-eight schools report that Ancient History is taught during the first semester of the first year, thirty-five during the second semester, and so on. One significant thing about these tables is the variability of practice. Although there is a fair concentration of frequencies, most of the subjects are reported as being taught in practically all of the four years. The number of schools not teaching the subjects is also significant. With the exception of American History which is not taught in four schools and Civics which is not taught in fifty-seven schools, no social science subjects can lay claim to being universally taught. Only five subjects—Ancient History, American History, Civics, Economics and Modern History—are reported as being taught in more than 50 percent of the schools. In addition to the variation of practice shown by these tables, the answers to the second question, which are not summarized in this report, indicate considerable leniency in admitting students from other years.

The tabulations of the answers to questions 3 and 4 are not reproduced here. They show in the first group of schools a median enrollment of from twenty to thirty-six, except for the last five subjects. In the second group of schools the median enrollment varies from thirteen to fifty-six. In the third group the variation is from twenty-six to one hundred thirty-eight, and in the fourth, from forty-two to two hundred thirty-eight. In general, the enrollments for the second semester are slightly less. One significant thing is the rather extreme variation in the enrollment for a given subject within

---

[1] 1. In what years and semesters are these subjects offered? 2. If pupils from other years are admitted, state the years. 3. How many pupils were enrolled last semester? 4. How many pupils are now enrolled?

TABLE I. SHOWING YEARS AND SEMESTERS IN WHICH VARIOUS SOCIAL SCIENCES ARE TAUGHT IN HIGH SCHOOLS

(Group I. Enrollment, 1-150. 122 High Schools Reporting)

Subject	Year and Semester								No Answer	Not Taught
	I		II		III		IV			
	1	2	1	2	1	2	1	2		
Ancient History	38	35	33	33	4	4	2	2	1	44
Ancient and Medieval History	6	7	18	19	3	3				87
Medieval History	1	2	7	4	5	2	1			108
Medieval and Modern History			25	24	25	25	3	3		67
Modern History	1	2	13	17	18	21	3	4		70
English History	1	1	1	1	7	7		1		109
General History	2	2	5	5	1	1				112
American History					33	14	74	37		3
Community Civics	17	11	1	3	3	2	1	6		84
Civics	1	2	1	2	7	25	26	64		13
Vocational Civics	4	4	3	2	1	1			1	115
Economics		1		3	17	8	29	18		56
Sociology				1	7	17	13	27		75
Social Science		1		1		2		1		120
Educational Guidance						2		3		119
Vocational Guidance	1		1	2	1	1	1		1	116

TABLE II. SHOWING YEARS AND SEMESTERS IN WHICH VARIOUS SOCIAL SCIENCES ARE TAUGHT IN HIGH SCHOOLS

(Group II. Enrollment 151-400. 197 High Schools Reporting)

Subject	Year and Semester								No Answer	Not Taught
	I		II		III		IV			
	1	2	1	2	1	2	1	2		
Ancient History	46	44	59	50	5	4	1	1		93
Ancient and Medieval History	31	35	56	56	3	4	2	3		121
Medieval History	2	7	4	13	4	1				171
Medieval and Modern History		1	37	34	42	48	1	1		115
Modern History	1	2	47	48	75	72	9	11		83
English History	2	1	2		15	16	4	2		175
General History	3	2	11	12	5	5	3	3		180
American History					39	24	166	105		0
Community Civics	41	38	9	10	2	5	1	6		122
Civics	4	3	6	3	28	40	51	116		26
Vocational Civics	4	6	1	2			1	1		182
Economics	1	1	4	4	31	41	50	60		68
Sociology	1	1	2	2	14	22	26	37		130
Social Science	1	2	1	1	1		5	6		185
Educational Guidance									3	194
Vocational Guidance	2	5	2	1			1			187

TABLE III. SHOWING YEARS AND SEMESTERS IN WHICH VARIOUS SOCIAL SCIENCES ARE TAUGHT IN HIGH SCHOOLS

(Group III. Enrollment, 401-800. 83 High Schools Reporting)

Subject	Year and Semester								No Answer	Not Taught
	I		II		III		IV			
	1	2	1	2	1	2	1	2		
Ancient History	22	17	21	17	1	1	1			42
Ancient and Medieval History	18	19	28	29	3	3				45
Medieval History	1	3	3	8	5	5	1	1		67
Medieval and Modern History		1	19	14	12	12	2	1		54
Modern History			21	21	32	31	2	2		32
English History	1	1	2	2	10	9	2			68
General History	1	1	7	7	1	1		1	1	77
American History			2	1	22	20	64	39		1
Community Civics	18	17	1	1			1	2		55
Civics			4	6	16	15	22	52		10
Vocational Civics	2	3		1					2	75
Economics			2	2	11	14	27	31		21
Sociology			2	1	2	8	10	12		58
Social Science	1	1					1	2	1	77
Educational Guidance	1						1		1	80
Vocational Guidance		2	2						1	78

TABLE IV. SHOWING YEARS AND SEMESTERS IN WHICH VARIOUS SOCIAL SCIENCES ARE TAUGHT IN HIGH SCHOOLS

(Group IV. Enrollment 801 and over. 73 High Schools Reporting)

Subject	Year and Semester								No Answer	Not Taught
	I		II		III		IV			
	1	2	1	2	1	2	1	2		
Ancient History	16	12	15	18	2	2	2	2	1	36
Ancient and Medieval History	16	16	16	16	3	3				39
Medieval History	1	7	5	3	9	3				51
Medieval and Modern History		1	7	8	9	8	1	1		52
Modern History			21	22	23	27	7	4		19
English History			3	4	14	17	2	2		52
General History			2	2	1	1				67
American History			1	1	13	13	54	49		0
Community Civics	17	11	4	4			3	3	1	40
Civics			4	4	11	9	35	45	1	8
Vocational Civics	4	2	3	1	1				1	65
Economics				1	11	12	30	37		14
Sociology		1			3	4	8	11	1	56
Social Science			5	3			4	3		62
Educational Guidance										73
Vocational Guidance	2	1	3	3			1	1		67

TABLE V. SHOWING YEARS AND SEMESTERS IN WHICH VARIOUS SOCIAL SCIENCES ARE TAUGHT IN HIGH SCHOOLS

(Total. 475 High Schools Reporting)

Subject	Year and Semester								No Answer	Not Taught
	I		II		III		IV			
	1	2	1	2	1	2	1	2		
Ancient History	122	108	128	118	12	11	6	5	2	215
Ancient and Medieval History	71	77	118	120	12	13	2	3		292
Medieval History	5	19	19	28	23	11	2	1		397
Medieval and Modern History		3	88	80	88	93	7	6		288
Modern History	2	4	102	108	148	151	21	21		204
English History	4	3	8	7	46	49	8	5		404
General History	6	5	25	26	8	8	3	4	1	436
American History			3	2	107	71	358	230		4
Community Civics	93	77	15	18	5	7	6	17	1	301
Civics	5	5	15	15	62	89	134	277	1	57
Vocational Civics	14	15	7	6	2	1	1	1	4	437
Economics	1	2	6	10	70	75	136	146		159
Sociology	1	2	4	4	26	51	57	87	1	319
Social Science	2	4	6	5	1	2	10	12	1	444
Educational Guidance	1					2	1	3	4	466
Vocational Guidance	5	8	8	6	1	1	3	1	2	448

a group of schools. Differences in the total enrollment must, of course, be kept in mind, but even so some of the variations seem of especial interest. For example, in Group I two schools reported an enrollment in Ancient History of between one hundred and one hundred twenty-five. Since no schools in this group have a total enrollment greater than one hundred fifty, all but a very few students in these two schools were studying Ancient History. Similar conditions for Ancient History as well as for other subjects are found in the other groups. This suggests that in many cases the course of study is in a state of change, or that distinctly different policies with reference to educational guidance are being followed in different schools.

The scope and content of the various social sciences. Questions[2] 7, 8, 12, 19 and 21 furnish information relative to the content

---

[2] 7. Does the class study the entire period covered by the text? 8. If not, what are the date limits? 12. Check the courses in which current events are taught? 19. Do you feel that more social science should be taught? 21. Give on line 1 for each subject: the author, the basic text, the date of its copyright, and the publisher.

TABLE VI. PERCENT OF HIGH SCHOOLS IN WHICH THE CLASSES STUDY THE ENTIRE PERIOD COVERED BY THE TEXT

(Based Upon Only Those Schools in Which the Subject is Taught)

Subject	Group I		Group II		Group III		Group IV		Total	
	Text Fol.	Text not Fol.	Text Fol.	Text not Fol.	Text Fol.	Text not Fol.	Text Fol.	Text not Fol.	Text Fol.	Text not Fol.
Ancient History	94	5	91	5	93	2	84	8	91	5
Ancient and Medieval	91	6	97	3	97	3	91	6	95	4
Medieval History	79	21	77	12	88	12	68	23	77	17
Medieval and Modern	96	4	93	3	90	3	86	5	93	4
Modern History	92	4	98	1	98	2	96	4	97	2
English History	84	8	86	5	86	7	95		89	4
General History	90		94	6	83		83		89	3
American History	96	3	90	7	92	4	93	6	93	5

TABLE VII. PERCENT OF HIGH SCHOOLS IN WHICH CURRENT EVENTS ARE TAUGHT

(Based Upon Only Those Schools in Which the Subject is Taught)

Subject	Percent of High Schools				
	Group I	Group II	Group III	Group IV	Total
Ancient History	30	45	54	49	43
Ancient and Medieval	40	47	55	47	48
Medieval History	29	38	50	55	44
Medieval and Modern	58	69	69	57	64
Modern History	60	78	75	78	74
English History	54	45	53	62	54
General History	40	47	67	17	44
American History	86	88	87	82	86
Community Civics	55	73	79	76	71
Civics	83	80	82	69	79
Vocational Civics	43	33	25	100	47
Economics	51	58	56	63	57
Sociology	53	52	44	47	51
Social Science	50	50	50	36	45
Educational Guidance	33		33		22
Vocational Guidance	17	10	20		11

of the various courses. The replies to question 7 given in Table VI indicate, except in the case of Medieval History, a very close adherence to the text books in regard to date limits.[3] The particular date limits for the various subjects will be noted in the next chapter devoted to the separate subjects. Table VII shows that current events are reported as being taught in a considerable percent of the classes in all subjects. As one might expect the percents are greatest for American History and for the various courses in Civics. A variety of texts is used, although there is considerable agreement upon a limited number. This subject also will be discussed further in Chapter III. The answers to question 19 and the general remarks made upon the questionnaire seem to indicate (1) that in the teaching of history less time should be spent on the political side and more on the economic, social and industrial phases, and (2) that in the teaching of the other social sciences the abstract and theoretical topics should be made more concrete and practical.

---

[3] In interpreting the facts in Table VI it is necessary to bear in mind that the percents have been computed on different bases in the case of the different subjects. For example, by reference to Table I we learn that Ancient History is taught in seventy-eight schools in the first group. Ancient and Medieval History is taught in thirty-five, Medieval in fourteen, Medieval and Modern History in fifty-five, etc. When it is stated in Table VI that 94 percent of the schools in Group I studied the entire period covered by the textbook in Ancient History, this simply means that this occurs in 94 percent of the seventy-eight schools in which Ancient History is taught. The 91 percent for Ancient and Medieval History is computed on thirty-five as a basis and the 79 percent for Medieval History is computed on fourteen as a basis.

The percents given in the remaining tables have been computed in the same way. Hence, it will be necessary to bear in mind this precaution in interpreting them.

Time devoted to social science courses. In a few schools which appear to be organized on a three-term basis as few as twelve weeks are given to certain of the social science studies. At the other extreme we find forty weeks being devoted to certain of the subjects. The number of minutes per week spent in recitation varies from 120 to 435. Usually a balance between the number of weeks devoted to the subject and the number of minutes per week is affected by requiring a class that spends relatively few minutes per week in recitation to carry a subject for a compensating longer period. For this reason the variations in time devoted to the different subjects is not as great as it appears from the above figures. The typical social science course extends through two semesters of eighteen weeks in length, and requires 225 minutes per week for recitation. Larger schools have a tendency to be in session longer than the smaller schools; therefore, they devote in most cases more time to the social science work.

The social sciences as required subjects. The replies to questions[4] 5 and 20 acquaint us with the status of the various social sciences as required subjects. Table VIII, which gives a summary of the replies to question 5, indicates a considerable difference of practice except in the case of American History and Civics. For the other subjects the predominant practice is to make them elective. The directions for replying to this question required an answer of either "yes" or "no". It is not unlikely that in a number of schools a subject is required for students pursuing certain courses but elective for others. As no opportunity was given to indicate this fact the number of affirmative replies probably was decreased. On the other hand, for the same reason an affirmative reply in some cases may have been slightly inaccurate. This character of the response required in this

---

[4] 5. Is this a required subject? 20. Check the courses that you feel the pupils should be required to take.

TABLE VIII. PERCENT OF HIGH SCHOOLS IN WHICH VARIOUS SOCIAL SCIENCES ARE REQUIRED

(Based Upon Only Those Schools in Which the Subject is Taught)

Subject	Group I		Group II		Group III		Group IV		Total	
	Req.	Not Req.	Req.	Not Req.	Req.	Not Req.	Req.	Not Req.	Req.	Not Req.
Ancient History	49	47	29	62	42	51	30	59	37	56
Ancient and Medieval	51	40	24	67	23	66	20	68	28	62
Medieval History	29	71	38	54	37	44	27	64	33	58
Medieval and Modern	33	58	24	67	24	66	29	52	27	63
Modern History	34	60	27	67	35	59	43	46	33	60
English History	23	62	9	91	13	67		90	10	80
General History	50	30	29	59	67	33	50	17	44	41
American History	85	11	85	11	79	17	89	6	85	11
Community Civics	50	37	46	43	46	36	46	33	47	39
Civics	79	14	72	22	71	26	66	19	73	20
Vocational Civics	30	70	33	53	38	50	75	25	42	50
Economics	45	47	16	74	24	70	13	68	23	67
Sociology	49	49	18	71	16	80	6	70	26	66
Social Science		50	17	83		67	27	46	16	65
Educational Guidance			34	66	33	67			22	33
Vocational Guidance	17	50	10	70	60	20	67	33	33	48

TABLE IX. PERCENT OF HIGH SCHOOLS WHICH FAVOR MAKING THE VARIOUS SOCIAL SCIENCE STUDIES REQUIRED
(Based Upon Only Those Schools in which the Subject is Taught)

Subject	Percent of High Schools				
	Group I	Group II	Group III	Group IV	Total
Ancient History	35	31	24	14	28
Ancient and Medieval	43	17	24	21	24
Medieval History	14	38	44	14	28
Medieval and Modern	44	34	17	14	32
Modern History	50	51	59	44	51
English History	31	36	7	14	23
General History	60	65	67	50	62
American History	88	86	87	86	87
Community Civics	74	67	61	79	70
Civics	81	82	77	69	79
Vocational Civics	14	40	25	38	32
Economics	59	61	42	37	53
Sociology	62	45	28	53	48
Social Science		42	33	55	42
Educational Guidance	67	33	67		56
Vocational Guidance	50	40	60	50	48

question should be borne in mind in interpreting Table VIII. In Table IX which summarized the replies to question 20 we find that in the opinion of the persons reporting some slight changes should be made with respect to requiring the different subjects. However, in general these changes were not marked. It is probably significant that in both Tables VIII and IX a decrease in the requirement of certain subjects is indicated as the size of the high school increased. In other subjects there is a little change and for some there is even an increase.

Permanency and success of the various social sciences. Questions[5] 6, 11, 17 and 18 give some indication of the permanency and the success of the various social sciences from the standpoint of pupils' interest. Table X, which summarizes the replies to question 6, indicates that, in the judgment of the persons reporting, these subjects are generally liked by a majority of students. Three-fourths or more of the schools replied affirmatively for most subjects. In a con-

---

[5] 6. Is this subject generally liked by the students? 11. Do you plan to offer this course again? 17. How long have these subjects been offered? 18. Check any courses that have proved unsuccessful.

TABLE X. PERCENT OF HIGH SCHOOLS IN WHICH THE VARIOUS SOCIAL SCIENCES ARE GENERALLY LIKED BY STUDENTS

(Based Upon Only Those Schools in Which the Subject is Taught)

Subject	Group I		Group II		Group III		Group IV		Total	
	Liked	Not Liked	Liked	Not Liked	Liked	Not Liked	Liked	Not Liked	Liked	Not Liked
Ancient History	63	24	75	16	71	19	68	8	70	18
Ancient and Medieval	68	26	84	13	76	13	65	9	76	14
Medieval History	79	21	73	19	69	12	64	14	70	17
Medieval and Modern	67	24	85	9	72	14	71	5	76	14
Modern History	83	15	77	11	84	12	66	4	77	11
English History	92		77	5	73	7	71	5	78	4
General History	70		82	6	83	17	50	17	74	8
American History	90	3	92	3	93	2	76	3	89	3
Community Civics	87		90	1	89		70	3	85	1
Civics	82	4	83	9	82	4	66	6	80	7
Vocational Civics	100		67		88		100		84	
Economics	83	8	84	8	82	8	71	5	81	7
Sociology	92	2	84	1	92		88	6	88	2
Social Science	100		75	8	67	16	73		74	7
Educational Guidance	100					33			33	11
Vocational Guidance	67		60		80		67		67	

TABLE XI. PERCENT OF HIGH SCHOOLS IN WHICH NOTEBOOK WORK IS REQUIRED FOR THE VARIOUS SOCIAL SCIENCES

(Based Upon Only Those Schools in Which the Subject is Taught)

Subject	Group I		Group II		Group III		Group IV		Total	
	Req.	Not Req.	Req.	Not Req.	Req.	Not Req.	Req.	Not Req.	Req.	Not Req.
Ancient History	81	17	74	17	76	19	81	16	78	17
Ancient and Medieval	77	20	80	19	74	26	68	20	77	19
Medieval History	86	7	65	23	81	19	77	14	75	17
Medieval and Modern	76	20	77	16	79	14	61	29	75	18
Modern History	75	23	75	20	86	14	74	19	76	19
English History	62	31	77	14	67	20	81	10	73	17
General History	20	40	71	23	67	33	67		56	26
American History	75	21	75	20	75	23	80	19	75	21
Community Civics	42	40	48	41	46	50	67	24	50	39
Civics	60	32	56	36	59	38	60	31	58	35
Vocational Civics	57	43	40	33	38	62	75	25	50	39
Economics	46	41	39	49	44	50	49	42	43	47
Sociology	55	41	40	39	36	56	35	65	44	45
Social Science	50	50	58	42	50	33	55	27	55	35
Educational Guidance	33	67	33	67		33			22	56
Vocational Guidance	33	17	50	40	40	40	100		56	26

TABLE XII. PERCENT OF HIGH SCHOOLS IN WHICH THE STEREOPTICON IS USED

(Based Upon Only Those Schools in Which the Subject is Taught)

Subject	Group I		Group II		Group III		Group IV		Total	
	Used	Not Used	Used	Not Used	Used	Not Used	Used	Not Used	Used	Not Used
Ancient History	26	68	22	64	34	63	89		35	56
Ancient and Medieval	23	74	24	70	11	87	35	62	23	73
Medieval History	22	71	23	62	44	56	50	41	35	56
Medieval and Modern	16	69	18	70	21	69	52	38	22	66
Modern History	29	67	19	70	22	78	35	57	25	68
English History	15	69	18	68	27	60	38	43	25	59
General History	10	60	47	53		67	33	17	28	51
American History	23	64	18	71	24	70	41	48	24	65
Community Civics	16	66	19	71	36	50	37	45	24	62
Civics	14	72	12	73	12	77	27	62	14	72
Vocational Civics	29	71	20	53	38	62	25	75	26	63
Economics	11	77	11	74	15	79	19	69	13	75
Sociology	21	70	15	66	12	80	12	76	16	71
Social Science		50	34	58	17	50	27	64	26	58
Educational Guidance	33	33				33			11	22
Vocational Guidance	17	33	10	80	20	60	16	67	15	63

TABLE XIII. PERCENT OF HIGH SCHOOLS IN WHICH THE SOCIALIZED RECITATION IS COMMONLY USED

(Based Upon Only Those Schools in Which the Subject is Taught)

Subject	Group I		Group II		Group III		Group IV		Total	
	Used	Not Used	Used	Not Used	Used	Not Used	Used	Not Used	Used	Not Used
Ancient History	46	42	37	47	46	51	24	73	38	50
Ancient and Medieval	43	49	40	54	50	42	41	53	43	50
Medieval History	50	50	35	46	44	44	23	73	36	54
Medieval and Modern	40	55	88	12	31	59	24	67	58	32
Modern History	50	46	43	49	50	40	41	54	45	48
English History	23	46	36	50	33	40	29	57	31	49
General History	30	50	23	65	17	50	33	50	25	57
American History	52	40	52	40	51	43	45	48	51	42
Community Civics	50	34	60	33	39	50	42	49	51	39
Civics	56	35	54	38	47	47	40	51	51	41
Vocational Civics	86	14	33	40	50	25	38	62	47	37
Economics	47	42	55	38	37	50	55	38	47	44
Sociology	58	36	54	43	60	36	65	35	57	39
Social Science	50	50	58	42	33	33	36	45	45	42
Educational Guidance	67	33		33		33			22	33
Vocational Guidance	17	33	50	40	40	40	33	67	37	44

TABLE XIV. PERCENT OF HIGH SCHOOLS IN WHICH SUPERVISED STUDY IS FREQUENTLY USED

(Based Upon Only Those Schools in Which the Subject is Taught)

Subject	Group I		Group II		Group III		Group IV		Total	
	Used	Not Used	Used	Not Used	Used	Not Used	Used	Not Used	Used	Not Used
Ancient History	45	45	36	61	46	51	33	62	40	55
Ancient and Medieval	46	51	47	50	47	42	29	62	44	51
Medieval History	43	57	27	58	44	44	55	45	41	51
Medieval and Modern	35	63	29	67	41	59	29	62	33	64
Modern History	44	54	37	58	43	55	35	61	39	57
English History	31	46	23	64	47	33	19	81	28	59
General History	50	50	29	65	50	50	16	67	36	59
American History	34	57	34	64	40	56	29	66	34	61
Community Civics	37	50	36	56	43	43	42	46	38	51
Civics	32	61	27	68	37	60	29	65	31	64
Vocational Civics	29	71	53	33	13	50	25	62	34	50
Economics	30	59	28	67	37	57	27	71	30	64
Sociology	41	55	27	67	48	40	35	65	35	59
Social Science	50	50	25	75	17	67	45	55	32	65
Educational Guidance	33	67				33			11	34
Vocational Guidance	17	33	10	90	20	60	50	33	22	59

siderable number of instances this question was not answered. The failure to do so probably means that the person replying to the questionnaire was very doubtful concerning the attitude of the students toward the subject.

Methods of instruction. Questions[6] 9, 10, 15 and 16 give certain information concerning the general procedure of instruction. The replies to these questions are summarized in Tables XI, XII, XIII and XIV. It is rather surprising to the writers that notebook work is reported as being required in such a large percent of the schools. As no attempt was made to define "notebook work" in the questionnaire blank the amount required in some of the schools probably is not large. The use of the stereopticon, as one might expect, is more frequently reported in the larger schools than in the smaller ones. Tables XIII and XIV give the summary for the socialized recitation and supervised study respectively. These phases of instructional procedure were not defined and the persons responding to the questionnaire no doubt differed in their interpretations. This naturally limits the significance of the replies. It is probably significant that the use of the socialized recitation does not appear to be restricted to any particular subjects, although, with the exception of Sociology, the highest relative frequencies are for American History and Civics. The percent of schools reporting the frequent use of supervised study is approximately one-third of the total number. It appears that supervised study occurs more often in smaller schools than in the larger ones.

---

[6] 9. Is notebook work required? 10. Is the stereopticon used? 15. Is supervised study used often? 16. Is the socialized recitation commonly used?

CHAPTER III.

SUMMARY BY SUBJECTS

Ancient History. Ancient history is reported as a separate course in more than 50 percent of the schools in each group. In addition, a considerable number of schools combine it with medieval history. It is slightly more frequently taught during the second year than the first, and in a few schools it is given during the third and fourth years. The enrollment compares favorably with that in the other social science subjects although it is usually exceeded by American History. The textbooks and the supplementary texts mentioned most frequently are given below:

Author	*Title*	*Frequency*
PRINCIPAL TEXTS		
Robinson and Breasted	History of Europe	49
West	Ancient World	39
Webster	Ancient History	32
Breasted	Ancient Times	30
Myers	Ancient History	24
Webster	Early European History	23
Betten and Kaufman	Ancient World	9
Ashley	Early European Civilization	9
West	Early Progress	8
SUPPLEMENTARY TEXTS		
Davis	Readings	26
West	Ancient World	6
Myers	Ancient History	6
Breasted	Ancient Times	6
Webster	Ancient History	5
	Total number of schools offering Ancient History	260

The answers to question 7 indicated that in approximately 90 percent of the schools offering Ancient History the textbook was followed with respect to date limits. When the date limits of the text are not adopted, 500 A. D. and 800 A. D. are most frequently mentioned. Only a few extend Ancient History to about 1000 A. D. Current events are reported as being taught in Ancient History courses in slightly more than two-fifths of the schools. More attention appears to be given to this topic in the larger schools than in the smaller ones. In more than one-third of the schools in which Ancient History is taught it is a required subject. This requirement is a little more frequent in the smaller schools, and the percent of replies indicating a belief that Ancient History should be required is somewhat less. Thus, it appears that if the persons replying to the questionnaire were able to assign Ancient History to the place which in their opinion it should occupy it would be required in fewer schools than at present.

Ancient History is reported as being generally liked by students in 70 percent of the schools offering it, with 12 percent not reporting. This is the smallest percent reported for any social science except "Educational Guidance" and "Vocational Guidance" which are taught in too few schools for the data to be significant. Apparently very few schools are planning to discontinue Ancient History as a separate course. Notebook work is required in about three-fourths of the schools. Use of the stereopticon is more frequent in Ancient History classes than in any other social science except Medieval History in which the same percent reports its use. The socialized recitation is used somewhat less frequently in Ancient History than in most of the other social science studies but in respect to supervised study it stands near the head of the list.

Ancient and Medieval History. This new combination of history is designed to take the place of the full year of Ancient History and one-half year of Medieval History in order that more time may be devoted to Modern History. About one-third of the schools reporting have adopted this combination in preference to the traditional one of Ancient History in the first or second year of the high school, followed immediately by Medieval and Modern History in the second or third year. The normal place for this combination subject is the second year, although it is occasionally reported for the third and fourth years. In schools having the same number of pupils the enrollment in this subject is reported as slightly larger than that in Ancient History.

Since the textbook defines the limits of the course in nearly every case the reader's attention is called to the list of textbooks and of supplementary texts that are most frequently used.

PRINCIPAL TEXTS

Author	*Title*	*Frequency*
Robinson and Breasted	History of Europe	72
Webster	Early European History	28
Robinson and Breasted	Outlines of European History	24
Breasted	Ancient Times	15
Robinson	Medieval and Modern Times	13
Ashley	Early European Civilization	8
SUPPLEMENTARY TEXTS		
Breasted	Ancient Times	6
Davis	Readings	5
Total number of schools offering Ancient and Medieval History		183

The final date limit of this course appears to be somewhere between 1476 and 1800, the year 1500 being given in most cases. In this combined course current events are taught more frequently than in the year course in Ancient History. More of the smaller schools than the larger require this subject. It has been offered in 124 schools for three years or longer, and in 28 schools for less than three years. Very few administrators plan to discontinue the course.

Medieval History. A critical examination of the data on this subject shows that Medieval History has probably been recorded separately in some cases when it should have been reported with either Ancient History or Modern History. However, no account has been taken of this error in the following discussion. Medieval History is taught most often during the second semester of the second year or the first semester of the third year. Usually pupils enrolled in the year above that for which the course is offered are admitted to it. In about 80 percent of the schools the textbook sets the limits of the course. However, it seems that as a rule the chronological period covered extends from the Germanic Invasions to about 1700. The textbooks used in this course are as follows:

PRINCIPAL TEXTS

Author	*Title*	*Frequency*
Robinson and Breasted	History of Europe	12
Robinson and Beard	Outlines of European History	12
Robinson	Medieval and Modern Times	11
Webster	Early European History	10
SUPPLEMENTARY TEXTS		
Robinson	Readings	5
West	Modern Progress	3
Hazen	(Not Mentioned)	2
Total number of schools offering Medieval History		78

Medieval History is required in about one-third of the schools in which it is taught. Usually these are the small schools which necessarily have a minimum of elective subjects. Hence, it is not surprising to find that slightly less than one-third of the persons replying to the questionnaire feel that this subject should be required. Except in the small schools it is not liked by the students as well as Ancient History. Only one school reported this course as unsuccessful and hence plans to discontinue it.

Medieval and Modern History. This subject is most frequently found in either the second or the third year of the high school. Only about one-third of the schools limit it to those pupils who have standing equivalent to the year in which it is offered. In most cases the smaller schools follow the textbook more strictly than the larger ones. About 27 percent of the schools replying to the questionnaire begin the work with the Decline of the Roman Empire, about 50 percent with the Germanic Invasions, and about 23 percent with the year 1000 A. D. More stress appears to be given current events than in any of the preceding subjects. The texts most frequently used are given below:

Author	*Title*	*Frequency*
	PRINCIPAL TEXTS	
West	Modern Progress	27
West	Modern World	19
Robinson and Beard	Outlines of European History	18
Harding	Medieval and Modern History	16
Myers	Medieval and Modern History	11
	SUPPLEMENTARY TEXTS	
Hazen	Europe Since 1815	12
Robinson	Readings	7
West	Modern Progress	5
Robinson and Beard	Outlines of European History	5
Total number of schools offering Medieval and Modern History		187

The replies indicate that the majority of the persons who filled in the questionnaire feel that this subject should be more universally required. About three-fourths of the schools report the subject as generally liked, about 7 percent state that it is unsuccessful, and about 4 percent expect to discontinue the course. Approximately 75 percent of the schools require notebook work, and the stereopticon is used in about 22 percent, the highest frequency being in the larger institutions. Supervised study and the socialized recitation are found in approximately one-third and three-fifths of these schools respectively.

Modern History. This subject is usually offered in the second or third year, the frequency being highest in the third year. Occasionally it is taught in the first or the fourth year. In one school practically all of the students are reported as being enrolled in the course. The textbook is more nearly followed in this subject than in any other of the social sciences. Those used most frequently are as follows:

Author	*Title*	*Frequency*
PRINCIPAL TEXTS		
Robinson and Beard	Outlines of European History	64
Robinson and Beard	Our Own Times	62
Hazen	Modern European History	28
West	Modern Progress	20
Robinson	Medieval and Modern Times	15
Ashley	Modern European Civilization	12
West	Modern World	12
Webster	Modern European History	10
SUPPLEMENTARY TEXTS		
Hazen	Modern European History	9
Robinson and Beard	Our Own Times	8
Robinson	Readings	6
	Total number of schools offering Modern History	271

In the smaller schools this course usually begins with the fifteenth century, but in the larger schools in most instances a later date is chosen. Current events are taught more frequently in connection with this course than with any other history course, except American History. It is interesting to note that about one-third of the schools require this subject for graduation, and that over one-half of the persons who answered the questionnaire think that it should be required. It has been offered for three years or longer in 190 schools, and for less than three years in 35 schools. It is more generally liked than Medieval and Modern History. Nine schools report it as unsatisfactory and six of these will not offer it again. Notebook work is required in about three-fourths of the schools. The data show that supervised study is in vogue in about 40 percent of these schools, and that the socialized recitation seems to be somewhat more widely used.

English History. This survival of the recommendation of the Committee of Seven is still found in over one-fourth of the high schools of the North Central Association. The usual year for this subject is the third, although sometimes it is offered in the second and occasionally in the first. When this subject occupies its normal position seniors usually are admitted and sophomores occasionally. No school reported that its pupils studied less than the entire period covered by the text. The texts most frequently used are as follows:

Author	*Title*	*Frequency*
	PRINCIPAL TEXTS	
Cheyney	Short History of England	40
Andrews	Short History of England	5
Cheyney	Readings	5
Larson	Short History of England	3
Walker	Essentials of English History	3
	SUPPLEMENTARY TEXTS	
Cheyney	Short History of England	2
Larson	Short History of England	2
Cheyney	Readings	2
Gardiner	History of England	2
	Total number of schools offering English History	71

Current events are taught in slightly more than 50 percent of the classes in English History. The smaller schools which offer this course report that it is liked best of all the social subjects. It is comparatively well liked in the larger schools. In three schools it was reported as unsuccessful.

General History. This subject is offered in less than 10 percent of the high schools of the North Central Association. It is found usually in the second year, although it was reported for all years. It is generally taught for the entire year and pupils from the year just above or below that in which the subject is scheduled are usually admitted. Only one school reported a failure to follow the work as outlined in the textbook. The textbooks which have the highest frequencies are as follows:

Author	*Title*	*Frequency*
	PRINCIPAL TEXTS	
Elson	Modern Times and the Living Past	10
Robinson-Breasted-Smith	General History of Europe	5
West	Modern Progress	5
Myers	General History	4
	SUPPLEMENTARY TEXTS	
Elson	Modern Times and the Living Past	3
West	Modern Progress	2
	Total number of schools offering General History	39

This course includes current events in more than 50 percent of the schools teaching it. It is interesting to note that in 44 percent of these schools the subject is required, and that the administrators in 62 percent of the schools think that it should be. This subject has been introduced into nineteen schools within the past three years. It is reported as being fairly well liked, and no schools pronounced it unsuccessful.

American History. American History is scheduled most frequently in the fourth year, although it is often found in the third. It is the only social science that is not offered by any school during the first year. Usually, pupils in the year above or below that in which the course is scheduled are admitted. When this course is a semester subject it is offered in the smaller schools most frequently during the first semester, but in the larger during the second semester. About 93 percent of the schools cover the entire period with which the text deals. It is probable also that they supplement the text with more recent history. The textbooks most frequently used are as follows:

Author	*Title*	*Frequency*
PRINCIPAL TEXTS		
Muzzey	American History	195
Fite	History of United States	47
Forman	Advanced American History	33
West	American History	30
McLaughlin	History of the American Nation	30
Hart	American History	22
SUPPLEMENTARY TEXTS		
Bassett	Short History of the United States	18
Muzzey	American History	14
Hart	New American History	7
	Total number of schools offering American History	471

Only one school did not report current events in connection with this course. American History is required in more than 80 percent of these schools. This is approximately the same as the percent of returns indicating that it should be a required subject. It is reported as being well liked in about 97 percent of the schools where it is taught, and as unsuccessful in slightly less than 3 percent of these schools. No school, however, expects to discontinue the course next year. Of these schools notebook work is required in 75 percent; the stereopticon is used in 24 percent; supervised study is found in 34 percent; and the socialized recitation finds a place in 51 percent. One significant and illuminating opinion, reported frequently, is that American History can not be done satisfactorily in one semester.

Community Civics. Community Civics is found most frequently in the first year and usually in the first semester. However, it is reported for every year of the high school. Few schools admit pupils from years other than that for which the course is scheduled. About 75 percent of the schools offering this course do the work as outlined by the texts. The texts in most frequent use are as follows:

PRINCIPAL TEXTS

Author	*Title*	*Frequency*
Hughes	Community Civics	96
Dunn	The Community and the Citizen	17
Hill	Community Life and Civic Problems	11
Hughes	Economic Civics	6
	SUPPLEMENTARY TEXTS	
Hughes	Community Civics	5
Hill	Community Life and Civic Problems	5
	Total number of schools offering Community Civics	174

Only two schools do not teach current events in connection with Community Civics. In at least one case this is due to the fact that current events are taught in connection with other courses, and this practice was resorted to in order to avoid duplication. Community Civics has been in the curriculums of eighty-two schools for three years or longer. During the past three years, however, it has been introduced into the curriculums of seventy-three schools. Only 2 percent of the schools report that the subject is not generally liked. Eight schools do not consider it successful and one plans its discontinuance. Notebook work is required in connection with this course in about 50 percent of these schools. The socialized recitation is used by an equal number. Supervised study is not found as frequently, and the use of the stereopticon is limited to about 25 percent of the schools offering the subject.

Civics. This course remains predominantly a fourth-year subject, although it is taught in a large number of schools during the third year. The data indicate that about 85 percent of all the schools offering courses in Civics follow the text rather literally. The textbooks most frequently used are as follows:

PRINCIPAL TEXTS

Author	*Title*	*Frequency*
Magruder	American Government	129
Guitteau	Government and Politics, Etc.	50
Ashley	New Civics	46
Woodburn and Moran	Citizen and Republic	39
Reed	Form and Functions of American Government	23
Garner	Government in the United States	18
Forman	American Democracy	18
Hughes	Community Civics	10
	SUPPLEMENTARY TEXTS	
Magruder	American Government	16
Forman	American Democracy	13
Bryce	American Commonwealth	9
Ashley	New Civics	7
Hughes	Community Civics	7
	Total number of schools offering Civics	418

In connection with Civics current events are taught in about 85 percent of the schools. The course is considered unsuccessful in twelve schools and reported as not being liked by the students in seven schools. This means either that it is generally liked in at least five schools reporting it as unsuccessful, or that the course was not properly estimated. Two schools plan its discontinuance. Notebook work is done in 58 percent of the schools. The socialized recitation is found nearly as frequently. Supervised study is found in about one-third of the schools, and the stereopticon is used for teaching purposes in only 14 percent, this low percent being due to the nature of the subject matter.

Vocational Civics. This subject is found most frequently in the first year of the high school course, but it is offered occasionally in every other year. Usually it is restricted to pupils who are in the year for which it is scheduled. As with most of the other social subjects the textbook furnishes the subject matter for the course. The texts most frequently used are as follows:

PRINCIPAL TEXTS

Author	*Title*	*Frequency*
Gowin and Wheatly	Occupations	13
Giles	Vocational Civics	11
Hughes	Economic Civics	2

(No supplementary texts were reported)

Total number of schools offering Vocational Civics 38

The persons replying to the questionnaire feel that Vocational Civics is given more prominence than it should have. This may be inferred from the fact that it is required in 42 percent of the schools but that only 32 percent of the replies expressed the opinion that it should be required. Vocational Civics has been taught in sixteen schools for three years or longer, and in seventeen schools for a shorter period of time. It is reported as being unsuccessful in six schools, and whether or not this is due to the nature of the course, it is interesting to note that no school intends to discontinue this subject. Notebook work is required and the socialized recitation is found in approximately 50 percent of the schools that offer this course. The stereopticon is used in one-half that number and supervised study in about two-thirds of that number.

Economics. Economics is offered most frequently in the fourth or the third year of the high school, yet six schools offer it in the first year. The data indicate that 82 percent of these courses are based predominantly upon some textbook. The texts used most frequently for this purpose are as follows:

PRINCIPAL TEXTS

Author	*Title*	*Frequency*
Thompson	Elementary Economics	104
Ely and Wicker	Elementary Principles of Economics	37
Carver	Elementary Economics	32
Burch and Nearing	American Economic Life	27
Bullock	Elements of Economics	19
Marshall and Lyon	Our Economic Organization	17
	SUPPLEMENTARY TEXTS	
Bullock	Elements of Economics	8
Carver	Elementary Economics	7
Ely	Outlines of Economics	6
Marshall and Lyon	Our Economics Organization	6
Seager	Economics	6

Total number of schools teaching Economics 316

This subject has been offered in 208 schools for three or more years and in 69 schools for a shorter period of time. Reports indicate that in about 7 percent of the schools the course is disliked by the pupils. Thirteen schools report it as unsuccessful and four of these schools will not offer it again. Notebook work is required in 58 percent of the schools, the stereopticon used only in about one-fourth that number, and supervised study and the socialized recitation found in 30 percent and 47 percent respectively.

Sociology. Sociology appears most frequently in the fourth year of the high school, yet a large number of schools offer it during the third year. Eight schools teach it regularly to freshmen. As a rule, however, only upper classmen are permitted to enter the course. It alternates in most cases with Civics or Economics. As a rule the textbook is the basis of the subject. The texts that are most frequently used are as follows:

Author	*Title*	*Frequency*
	PRINCIPAL TEXTS	
Towne	Social Problems	67
Ellwood	Sociology and Modern Social Problems	46
Burch and Patterson	American Social Life	22
Tufts	Real Business of Living	4
	SUPPLEMENTARY TEXTS	
Ellwood	Sociology and Modern Social Problems	9
Towne	Social Problems	4
Burch and Patterson	American Social Life	4
Tufts	Real Business of Living	2
	Total number of schools teaching Sociology	156

Practically one-half of these schools teach current events in connection with this course. About 25 percent of the schools offering Sociology require it for graduation. The persons answering the questionnaire in about 50 percent of the schools believe the subject should be required. Hence, again theory and practice differ. Indications are clear that Sociology is fast coming into the high school curriculum. Its presence is reported by fifty-nine schools for three years or longer, and by seventy-five schools as having been introduced within the past three years. It is reported as unsuccessful in seven schools and as not being generally liked in three schools. It will not be offered next year by three schools. Notebook work is required in 44 percent of the schools, the stereopticon is found in 16 percent, and supervised study and the socialized recitation are found in 35 percent and 57 percent, respectively.

Social Science. This course is offered most frequently to seniors or to sophomores. In nearly every instance it is limited to the students of the class for which it is offered. About three-fourths of the schools giving the subject adhere strictly to the textbook. The textbooks that are most frequently used are as follows:

PRINCIPAL TEXTS

Author	*Title*	*Frequency*
Hill	Community Life and Civic Problems	4
Burch and Patterson	American Social Life	4
Tufts	Real Business of Living	3
Towne	Social Problems	2
	Total number of schools offering Social Science	31

Current events are taught in this course in about three-fourths of the schools. About 16 percent require the subject for graduation. Indications from the opinions of those persons answering the questionnaire are that the subject should be required much more frequently that it is at present. Social Science has been offered for three years or longer by seven schools, and for less than three years by nineteen schools. It is reported as not being generally liked in two schools, and as unsuccessful in one school which plans to discontinue it. Notebook work is required in 55 percent of the schools offering the subject, the stereopticon is used in 26 percent, supervised study in 32 percent, and the socialized recitation in 45 percent of these schools.

Educational Guidance. This subject is offered in a very few schools, and appears either in the first semester or in the last three semesters of the high school course. When offered in the first year it is likely that one phase of its work is related directly to the students' future work in the high school. This subject does not appear to have a very definite content, and probably the subject matter of the course usually depends upon the interests of the instructor. In several cases "How to Study" by Sandwick is mentioned as a basic text. Current events are taught in almost one-half of these courses.

Of the schools teaching Educational Guidance 20 percent require it for graduation. However, 56 percent of the persons who answered the questionnaire think that it ought to be required. It is not generally liked in one of the five schools reporting its presence. Notebook work and the socialized recitation are used in 20 percent of these schools.

Vocational Guidance. This subject is offered in either the second or the first year of the high school course. In nearly every case the classes are restricted to students with standing equivalent to the year in which the course is scheduled. "Occupations" by Gowin and Wheatley is used almost universally as a text. No other text was mentioned more than once. One-third of the schools require this course for graduation but the conclusion based on the answers to the questionnaire is that it should be required to a much larger extent. This subject also has made its appearance within the last few years, although only nine schools report that it has been taught for less than three years. It is generally liked in all the schools represented, is reported successful, and will be offered again in every case. Notebook work is required in 50 percent of the schools. The stereopticon is used in 15 percent, and supervised study and the socialized recitation reported in 22 percent and 37 percent, respectively.

BULLETIN NO. 14

BUREAU OF EDUCATIONAL RESEARCH
COLLEGE OF EDUCATION

THE USE OF DIFFERENT TYPES OF THOUGHT QUESTIONS IN SECONDARY SCHOOLS AND THEIR RELATIVE DIFFICULTY FOR STUDENTS

by

WALTER S. MONROE,
Director, Bureau of Educational Research

and

RALPH E. CARTER
Associate Professor of Educational Psychology
Indiana University, Extension Division

PRICE 30 CENTS

PUBLISHED BY THE UNIVERSITY OF ILLINOIS, URBANA
1923

UNIVERSITY
of ILLINOIS
PRESS

PREFACE

This bulletin is a report of one phase of a larger investigation relating to the study habits of school children. The types of questions asked by teachers of their students are important both because of the mental processes which occur in answering them and because an intimate relation exists between the questions asked and the detailed objectives toward which the students work.

Mr. Carter was consulted in the preparation of the questionnaire. He rendered valuable assistance in deciding upon the tentative list of types of thought questions. Later, advantage was taken of his presence at the University of Illinois during the Summer Session of 1922 to secure assistance in the preparation of this report. He is largely responsible for its general organization.

WALTER S. MONROE, *Director.*

February 1, 1923

The Use of Different Types of Thought Questions in Secondary Schools and Their Relative Difficulty for Students

Recent emphasis upon questioning in instruction. The first quantitative study of the questions which teachers use in the classroom was made by Miss Romiett Stevens.[1] In this analysis of questions a distinction was made between those that required memory only and those that required thought. Miss Stevens' discussion has been widely quoted and has been very influential in causing teachers to give more attention to the types of questions asked of their students. It is altogether likely that this investigation is primarily responsible for a material increase in the percent of thought questions which are asked in our schools today. Many of the most commonly used texts on methods of teaching devote separate space to the topic of questioning. Its importance has been especially emphasized by Charters,[2] Strayer,[3] Parker,[4] and Colvin.[5]

Analysis of thought questions. An examination of typical thought questions from the standpoint of the mental processes which they require in answering reveals certain significant differences. Some questions require the student to "compare" two or more ideas; others ask that he "summarize;" still others demand that he "give reasons why." The probable mental processes occurring in the reflective thinking called for rather than the form of the question or the wording of the answer have been made the basis of these types of thought questions. Altho there is doubtless considerable overlapping in the types of mental processes which usually occur in answering such questions, there appears to be sufficient differences to justify the recognition of a number of types of questions. In this investigation the following tentative list of types of thought questions was used. A more refined analysis would yield probably some additional types, but, on the other hand, for practical purposes it is possible that some combinations of types might profitably be made. It is recognized that the mental processes of different students doubtless vary in answering the same question. In fact a thought question for one student may be a memory question for another student, or even for the same student on the following day.

---

[1] Stevens, Romiett. "The question as a measure of efficiency in instruction," Teachers College Contributions to Education No. 48. New York: Teachers College, Columbia University, 1912.

[2] Charters, W. W. Methods of Teaching. Chicago: Rowe Peterson and Company, 1912, p. 444. (Chapter XVII).

[3] Strayer, G. D. A Brief Course in the Teaching Process. New York: Macmillan Company, 1912, p. 315. (Chapter XI)

[4] Parker, S. C. Methods of Teaching in High Schools. Boston: Ginn and Company, 1915, p. XXV, 529. (Chapter XX)

[5] Colvin, S. S. Introduction to High School Teaching. New York: Macmillan Company, 1917, p. XXI, 451. (Chapter XV)

1. **Selective recall**—basis given.

   Name the presidents of the United States who had been in military life before their election.

   What do New Zealand and Australia sell in Europe that may interfere with our market?

2. **Evaluating recall**—basis given.

   Which do you consider the three most important American inventions in the nineteenth century from the standpoint of expansion and growth of transportation?

   Name the three statesmen who have had the greatest influence on economic legislation in the United States.

3. **Comparison of two things**—on a single designated basis.

   Compare Eliot and Thackeray in ability in character delineation.

   Compare the armies of the North and South in the Civil War as to leadership.

4. **Comparison of two things**—in general.

   Compare the early settlers of the Massachusetts Colony with those of the Virginia Colony.

   Contrast the life of Silas Marner in Raveloe with his life in Lantern Yard.

5. **Decision**—for or against.

   Whom do you admire more, Washington or Lincoln?

   In which in your opinion can you do better, oral or written examinations?

6. **Causes or effects.**

   Why has the Senate become a much more powerful body than the House of Representatives?

   What caused Silas Marner to change from what he was in Lantern Yard to what he was in Raveloe?

7. **Explanation** of the use or exact meaning of some phrase or statement in a passage.

8. **Summary** of some unit of the text or of some article read.

9. **Analysis.** (The word itself is seldom involved in the question.)

   What characteristics of Silas Marner make you understand why Raveloe people were suspicious of him?

   Mention several qualities of leadership.

10. **Statement of relationships.**

    Why is a knowledge of Botany helpful in studying agriculture?

    Tell the relation of exercise to good health.

11. **Illustrations or examples** (your own) of principles in science, construction in language, etc.

12. **Classification.** (Usually the converse of No. 11).

    What is the principle involved here? What is the construction? To what class or genus does this individual belong?

13. **Application** of rules or principles in new situations.

14. **Discussion.**

    Discuss the Monroe Doctrine.

    Discuss early American Literature.

15. **Statement of aim**—author's purpose in his selection or organization of material.

    What was the purpose of introducing this incident?

    Why did he discuss this before that?

16. **Criticism**—as to the adequacy, correctness, or relevancy of a printed statement, or a classmate's answer to a question on the lesson.

17. **Outline.**

18. **Reorganization of facts.** (A good type of review question to give training in organization.)

    The student is asked for reports where facts from different organizations are arranged on an entirely new basis.

19. **Formulation of new questions**—Problems and questions raised.

    What question came to your mind?

    What else must be known in order to understand the matter under consideration?

20. **New methods of procedure.**

    Suggest a plan for proving the truth or falsity of some hypothesis.

    How would you change the plot in order to produce a certain different effect?

Relation to educational objectives of types of questions asked by teachers. Incidentally it may be noted that the types of questions used both for stimulating and directing the mental activity of the learner and for measuring the results of teaching reflect in a subtle way the educational objectives of a teacher. The objectives thus indicated may not agree with those stated by the teacher but they nevertheless are an index of the objectives toward which the students direct their efforts. In studying, students tend to prepare to answer the kind of questions which they think the teacher will ask. It is altogether likely that the study objectives of students are influenced more by the kind of questions asked than by direct statements of aims in the course.

Purpose and method of this investigation. The purpose of this study, which is Sub-project II of our investigation of the study habits of high school pupils, was announced as follows: "To determine the extent of the use of different types of thought questions in actual school practise and the relative difficulty of these types for students." The method employed was to submit a questionnaire to a large number of teachers in the seventh and eighth grades and in the high school. In this questionnaire the list of the twenty types of thought questions given on pages 6-7 was reproduced. The teachers were urged to study these types carefully before answering any of the questions. It was pointed out that an attempt to classify a question according to this plan might lead to the conclusion that it is a combination of two or three types. However, in such a case, it is probably true that part of the work of answering has been done by the author in the textbook, leaving to the student only that phase of the question which would definitely classify it as belonging to one of the above types.

Plan of summarizing data. Usable replies were received from 199 teachers representing almost every school subject and all parts of the state. A few of these replies did not give answers to one or more of the questions but they were included in the tabulation for the other questions. By subject the number of replies were as follows: English, 41; History, 48; Science, 41; Mathematics, 31; Foreign Language, 12; Geography, 11; Commercial subjects, 6; Agriculture, 4. Taking only the replies from teachers of English, History, and Science there were 26 for grades VII and VIII combined; 34 for grades IX and X; and 45 for grades XI and XII. It was thought advisable to summarize separately the replies for English, History and Science. Those received from teachers of all other subjects were grouped together. The replies to Questions I and IV obviously do not lend themselves to statistical treatment. The data yielded by the other three questions are summarized in the following tables.

Limitations of replies to the questionnaire. Several teachers took occasion to mention difficulties which they experienced in filling out the questionnaire blank. That some difficulties were encountered is not surprising because the concept of different types of thought questions was undoubtedly new to most, if not all, of the teachers. The limitations of space prevented an extensive description of each type. Undoubtedly teachers differed in their interpretations of the various types. Furthermore, in answering Question II they were asked not only to differentiate between thought questions and memory questions but also to classify the thought questions under the various types. In doing this, they had to rely upon their memory of the questions they had asked. For these reasons the replies to Question II must be thought of as giving only a very rough indication of the practise of teachers with reference to the types of questions which they asked. A few teachers reported that their replies were based on a careful analysis of lists of questions which have been used and preserved. If these lists were reasonably complete their replies should be more accurate than those received from teachers who relied upon their memory of the questions they had asked.

It may be pointed out that other methods of ascertaining the relative frequency of use of different types of thought questions would not eliminate all difficulties. For example, if a trained investigator should visit the classes of 199 teachers in various parts of the state a sufficient number of times to obtain a fair sample of the habitual practise and should try to make a record of all types of questions asked he would encounter difficulties in definitely classifying them. It is not possible to judge accurately of the mental processes stimulated by a question unless one is acquainted with the previous experiences of the pupils in the field of this question. This has been referred to already in pointing out that what constitutes a thought question for one pupil may be merely a memory question for another. The analysis of stenographic reports of lessons would present certain difficulties for the same reason. The teacher knows better than any one else what acquaintance a student has with a topic and what the textbook states about it. For this reason he has the advantage of a casual observer.

Additional types of questions mentioned. Question I was asked in order to obtain suggestions for supplementing the list of type questions for further study in this field and also to give each teacher an opportunity to report fully his own practise in the use of different types of thought questions in case he did not consider the list adequate. The replies to the questionnaire show that the number of teachers who did not report the use of all of the types given in the list

UNIVERSITY OF ILLINOIS
BUREAU OF EDUCATIONAL RESEARCH
COLLEGE OF EDUCATION

Types of Thought Questions

Name............................Address..
School Subject..........................School grade.......................

I. Write in on the blanks at the bottom of the list below any additional types of thought questions you use to any great extent in the grade and subject you have chosen. (Include these additional types in answering all other questions.)

II. What percent of *all* of your questions to the class are of each type listed? (See directions, p. 11). Put the percent of each type in the column on the right.

Types of Thought Questions	Percent of all questions
1. Selective recall—basis given..............................	....
2. Evaluating recall—basis given..........................	....
3. Comparison of two things—on a single designated basis. ...	....
4. Comparison of two things—in general.....................	....
5. Decision—for or against..................................	....
6. Causes or effects...	....
7. Explanation...	....
8. Summary...	....
9. Analysis..	....
10. Statement of relationships..............................	....
11. Illustration or examples................................	....
12. Classification..	....
13. Application...	....
14. Discussion..	....
15. Statement of aim..	....
16. Criticism...	....
17. Outline...	....
18. Reorganization of facts.................................	....
19. Formulation of new questions............................	....
20. New methods of procedure................................	....
21.	....
22.	....
23.	....
24.	....

III. In the left hand margin make an "X" opposite each of the five types for which students answers are least likely to be satisfactory. (See directions, p. 11)

IV. Choose three of the types you marked with an "X." State for each some of the students' most common faults in procedure in answering it. Mention faults of omisson as well as faults of commission. Mention only faults peculiar to this type of question. Use other side of this sheet if necessary. (See directions below.)

V. If we consider all unsatisfactory answers made by students in school work, what percent of them, in your opinion, are due primarily to bad habits of procedure in answering questions? Give separate replies for memory questions ....% and thought questions............%

The following detailed directions for answering the questionnaire were given:

Before answering any of the questions, decide what subject and what grade (any from VII to XII) you will have in mind in all of your replies. Answer the questions with reference to only *one* school subject. Fill in the blanks at the top of the first page of the questionnaire.

Question I. Examine the list of types given above to see if there are not some other types of thought questions that you use to a considerable extent. Altho the given list may seem long and inclusive at first, it is not complete.

Question II. Before you put down your estimates for Question II, you should decide what is your proportion of thought questions and of pure memory questions. This question calls for a further analysis of your thought questions. It may be advisable to make rough estimates for all before you fill in the data on the questionnaire. After you have the differences between the types in mind, make the best estimate you can, even tho you may not be very sure in many cases. You are merely declaring your best judgment, not guaranteeing that it is infallible. *We want your opinion of what your practise is, not of what you think it ought to be.*

Question III. In Question III, we are concerned with the *process* of answering so we must assume that the student has the information necessary for a satisfactory answer if he will only use it as the type of the question requires. The common starting point for an informal diagnosis of a student's study faults is in his unsatisfactory answers. An unsatisfactory answer may indicate lack of information or certain faulty habits in answering questions. There are many cases in which teachers find out by other means that the student has enough information to give a satisfactory answer but has failed because he did not take the necessary steps and precautions in replying to the question. When some students correct examination papers returned to them, they are heard to say, "I knew that, but I didn't think of it," or "I knew that but I didn't know that was what the question meant." Some of the faults in procedure are common to all kinds of questions; others are peculiar to particular types. In Question III, you are asked to decide on the difficulty of the types for students. In other words, in which type do they have the worst habits of procedure?

Question IV. It will be helpful in answering Question IV to think of some particular questions of the type under consideration and then express the students' faults in fairly general terms.

Question V. As long as the unsatisfactory answer is a resultant of poor methods of preparation and poor habits of answering questions, we must correct the latter or take it into consideration in inferring what methods of study are needed. It is important to get an estimate from a large number of teachers of the percent of failures (in answers) that are due primarily to bad habits of procedure in answering questions.

is very much greater than the number mentioning additional types. Only thirty-nine teachers (19.6 percent) indicated that they used all twenty types of questions. Thirteen teachers supplemented the list, each writing in from one to three additional types. These teachers were distributed among the different subjects as follows: seven in History, three in English, two in Science, and one in Mathematics. Only four of those suggesting additional types of questions indicated that they used all twenty of the types given in the questionnaire list.

Some of the additional types of questions suggested are clearly included in the list given in the questionnaire. The following are typical: (A) "Connections between historical events," Type 4 or 6; (B) "Cause and results," included in Type 6; (C) "Application to the present time," included in Type 13; (D) "Determining the significant word in a statement," included in Type 19. Other suggested types are not so clearly included but several of them under a fairly loose interpretation may be placed with the types listed. Illustrations of these are: (A) "Imagining the results if conditions were different," a special case of Type 6; (B) "Questions on appreciation such as, which do you like best?" a fairly common form of Type 2; (C) "Identify a known character in a scene in the story when his name is not mentioned by the author," a special case of Type 12; (D) "What should you judge from these facts?" This last is broad enough to cover many types. Frequently, it would fall under Type 6. There were a few suggestive questions which are not so easily classified under the 20 types given. Some of these are: (A) "Trace the development," mentioned by two teachers; (B) "Estimate the importance of"; and (C) "Why is this statement true?" a very common question in geometry. By a very liberal interpretation these might be put under Types 9, 2, and 6 respectively but it is likely that many teachers would consider them sufficiently different to justify naming them as additional types.

The answers to the first question indicate that the list of types is reasonably complete. For practical purposes it is probably better to give a rather loose interpretation in classifying special cases than to try to extend the list. Largely for this reason the authors decided to limit the summary of the replies to the other parts of the questionnaire to the original twenty types.

Frequency of use of different types of questions. Question II of the questionnaire was asked in order to secure answers for the following:

1. Which types of questions are most commonly used by teachers in classroom instruction?
2. Which types of questions are most characteristic of the instruction in different subjects?
3. Do teachers in the more advanced grades tend to ask different types of questions from those in the lower grades?

Table I gives a detailed summary of the replies by forty-eight teachers of History. This table shows that four of these teachers did not report any use of questions requiring selective recall (Type 1); one teacher reported that this type of question formed 1 percent of all his questions; five teachers indicated that this type made up 2 percent of all the questions which they asked; eleven named 5 percent as an index of its frequency; eight teachers considered that 15 percent of their questions were of this type; and another eight teachers indicated that more than 15 percent of their questions called for selective recall. The median practise is 7.5 percent.

The outstanding characteristic of the table is the variability of practise which it indicates. In the case of each type of question there were one or more teachers who gave no indication of its use. For several of the types the number of teachers not indicating any use is surprisingly large. Furthermore, there is no type of question which did not receive an indication of at least 5 percent by one teacher. Most of the types received an indication of 10 percent or more. Thus, if we may assume that the replies to this portion of the questionnaire are a reasonably true indication of practise, it is clear that teachers of History vary widely in the types of thought questions which they ask of their pupils and hence necessarily vary widely in the detailed objectives which their pupils strive to attain.

Similar tabulations were made for English, Science and other subjects combined. The median frequencies for the four groups of subjects are given in Table II. If we consider only English, History and Science, students are most commonly required to answer the following types of thought questions: cause and effect, Type 6; selective recall, Type 1; discussion, Type 14; and evaluating recall, Type 2. The types of questions which are asked least frequently are: formulation of new questions, Type 19; new methods of procedure, Type 20; and reorganization of facts, Type 18.

Different subjects require different mental processes. An examination of Table II reveals that the frequency of use of some of the

TABLE I. TYPES OF THOUGHT QUESTIONS ASKED BY 48 TEACHERS OF HISTORY

Types of Questions	Percent of all Thought Questions																
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Over 15	Median
1. Selective recall	4	1	5	2		11		2		1	6					8	8	7.5
2. Evaluating recall	2	1	4	4	3	7	2	3	3	2	5		3			6	3	7.3
3. Comparison—single	6	5	4	1	4	16			2	2	5		1		1	1		5.3
4. Comparison—general	4	1	8	3	2	13	2		4	1	6					2		5.5
5. Decision	10	11	5	4	3	10		1	1		2					1		2.6
6. Causes or effects	1		1	1	3	6	2	2	3		12		1			4	10	10.4
7. Explanation	15	4	8	8	1	5	1				3					2		2.6
8. Summary	5	1	7	1	2	15			1	1	10					1	3	5.5
9. Analysis	7	7	8	4	5	11		1	1	1	2							3.5
10. Relationships	13	4	7	4	5	4	4	1	4		2							3.0
11. Illustration or examples	25	5	3	2	4	6					1		1			1		1.0
12. Classification	26	4	6	2	2	6	1		1								1	.9
13. Application	24	5	7	3	3	2		1			2		1					1.0
14. Discussion	4		1		1	6	2	3	3	1	12					6	9	10.3
15. Statement of aim	24	7	6	3	4	2										1		1.0
16. Criticism	12	5	12	4	4	6	1	1	1		2							2.6
17. Outline	6	3	5	4	4	11	1				9				2		1	5.2
18. Reorganization of facts	14	8	8	6		5			2		4					1		2.3
19. Formulation of new questions	31	3	3	7	3				1		1							.8
20. New methods	24	5	9	2	3	3			1		1							1.0

TABLE II. MEDIAN FREQUENCY OF USE OF DIFFERENT TYPES OF THOUGHT QUESTIONS IN ENGLISH, HISTORY, SCIENCE AND OTHER SUBJECTS

Types of Questions	Median percent of Use			
	English	History	Science	Others
1. Selective recall	5.4	7.5	5.3	5.1
2. Evaluating recall	5.9	7.3	2.8	.96
3. Comparison—single	5.4	5.3	3.9	1.6
4. Comparison—general	5.0	5.5	3.1	1.9
5. Decision—for or against	4.4	2.6	1.1	1.2
6. Causes or effects	5.8	10.4	10.2	4.8
7. Explanation	10.0	2.6	3.3	5.1
8. Summary	5.6	5.5	2.9	2.1
9. Analysis	5.7	3.5	1.8	3.1
10. Relationships	1.8	3.0	5.7	3.4
11. Illustration or examples	3.6	.96	8.6	5.1
12. Classification	1.6	.9	5.2	5.1
13. Application	2.9	1.0	5.1	5.5
14. Discussion	5.7	10.3	5.3	1.1
15. Statement of aim	4.4	1.0	.7	.8
16. Criticism	5.2	2.6	2.9	3.9
17. Outline	3.8	5.2	.98	.8
18. Reorganization of facts	2.1	2.3	.98	.9
19. Formulation of new questions	1.9	.8	1.7	1.8
20. New methods	2.1	1.0	1.5	3.6

types of thought questions varies in the different subjects. For example, Type 7 "Explanation of the use or exact meaning of some phrase or statement in a passage" is the most frequently used type of question in English but it is eleventh in frequency of use in History. Asking students to give an illustration or example, Type 11, is next to the most frequent type of question in Science but it is seldom used in History. Type 10, which asks the student to state a relationship, is frequently used in Science but is infrequently asked of students in English. Summaries and outlines are required more frequently in History and English than in the other subjects. Thus, if our assumption that different types of questions require different mental processes is valid, we have here evidence that different types of mental processes are required of students in the different subjects. Hence, we should expect to find that the problem of directing the learning of students is different in different subjects.

TABLE III. PERCENT OF TEACHERS REPORTING NO USE OF CERTAIN TYPES OF THOUGHT QUESTIONS

Types of Questions	Grades VII-VIII	Grades IX-X	Grades XI-XII
1. Selective recall	3	18	11
2. Evaluating recall	6	29	13
3. Comparison—single	18	9	18
4. Comparison—general	9	18	13
5. Decision—for or against	21	32	27
6. Causes or effects	9	6	4
7. Explanation	23	26	18
8. Summary	6	23	22
9. Analysis	21	23	29
10. Relationships	35	18	20
11. Illustration or examples	41	26	16
12. Classification	50	29	24
13. Application	46	35	22
14. Discussion	9	15	13
15. Statement of aim	63	46	42
16. Criticism	26	23	24
17. Outline	26	21	31
18. Reorganization of facts	35	41	42
19. Formulation of new questions	68	44	44
20. New methods	44	53	32

Variations in the use of type questions in different school grades. The replies of 113 teachers have been summarized in Table III on the basis of the school grades in which they were teaching. The subjects involved are English, History, Science and Geography. Instead of giving the median frequency of use, the percent of teachers indicating no use of the given type of question has been calculated. Thus, 3 percent of the teachers in the seventh and eighth grades reported no use of selective recall. In the ninth and tenth grades 18 percent of the teachers did not use this type and in the eleventh and twelfth grades, 11 percent. The general impression prevails that students in advanced classes are asked more difficult types of questions than students in the lower grades. A comparison of the data for the three grade groups shows that the differences are not very great. In general, it appears that the variations in the use of thought questions are greater for different subjects than for the different grades.

The relative difficulty of the different types of questions for students. In Question III teachers were asked to designate the five

TABLE IV. TYPES OF QUESTIONS FOR WHICH STUDENTS' ANSWERS ARE LEAST SATISFACTORY

Types of Questions	English	History	Science	Others	Total	Rank
1. Selective recall........	5	10	8	14	37	15
2. Evaluating recall......	12	12	11	10	45	8
3. Comparison—single....	6	7	9	5	27	19
4. Comparison—general..	12	12	9	14	47	6
5. Decision—for or against	1	9	2	6	18	20
6. Causes or effects.......	17	17	19	24	77	1
7. Explanation..........	15	12	15	25	67	3
8. Summary............	10	14	9	14	47	6
9. Analysis..............	13	15	11	21	60	5
10. Relationships.........	5	17	10	20	42	10
11. Illustration or examples	5	4	10	14	33	17
12. Classification..........	5	4	9	16	34	16
13. Application...........	16	12	18	29	77	1
14. Discussion............	10	14	5	10	39	13
15. Statement of aim......	15	12	5	10	42	10
16. Criticism.............	8	14	7	10	39	13
17. Outline...............	6	12	6	9	33	17
18. Reorganization of facts.	12	19	13	19	63	4
19. Formulation of new questions...........	9	6	8	18	41	12
20. New methods.........	6	7	7	24	44	9

types of questions in which "students' answers were least likely to be satisfactory." A summary of their replies is given in Table IV. The first line of this table should read as follows: Questions calling for selective recall were named among the five most difficult types by five teachers of English, ten teachers of History, eight teachers of Science, and fourteen teachers of other subjects. This type of question ranks fifteenth in the frequency of mention among the five most difficult types. The two types most frequently mentioned were Type 6, causes and effects, and Type 13, application. Both of these were mentioned by seventy-seven teachers or 38.7 percent. Altho we may say that in the judgments of those answering the questionnaire these two types are the most difficult for students, it should be noted that over 60 percent of the teachers did not list them among the five most difficult types.

Here, as in other tables, the variation in the replies of teachers is very extreme. Even when we consider the teachers of a single subject we find marked variations in their judgments concerning the difficulty of the different types of questions. No type of question is so easy that it is not included among the most difficult five by several teachers. This variation in judgment is probably due largely to the fact that teachers have not recognized the distinction in types of questions which are made here and, furthermore, they have not analyzed the responses of their students in order to ascertain the nature of the difficulties which the students encounter in answering the questions.

Common faults of procedure in answering different types of questions. In Question IV the teachers were asked to select three of the types which they considered most difficult and to state for each of these some of the students' most common faults in procedure in answering them. An examination of the replies to this question suggests that a considerable number of teachers are satisfied with telling students that the answer is wrong, or what the answer should have been, and that they fail to give much thought to the students' faults of procedure in answering the question. It may be pointed out that questions are only a means to an end and this end is the education of the student. The answer is in itself relatively unimportant. The question fulfils its function only when it stimulates educative processes in the mind of the student. If the mental processes which the question initiates are not educative the asking of the question has been largely futile. Hence, it becomes highly important for the teacher to give attention to the procedure which the student employs in answering questions in order that the faults of procedure may be corrected so that the student's mental processes will become most effective in his education.

Of the 199 teachers from whom replies to the questionnaire were received, twenty-three did not attempt to answer Question IV at all and twenty-four others discussed all three types of questions together or stopped after discussing only one or two. Furthermore, a number of other teachers gave answers which show that they failed to grasp the significance of this question. In substance they said that the difficulty in students' procedure in answering a question is inability to do what is called for. For example, one said that the fault in answering Type 2, evaluating recall, is that "pupils are not always able to evaluate." Another said that in questions on aim "they seem unable to see any motive behind the statement of the author." Still another said that the fault in Type 6, causes and effects, is that "they are unable to see causes and effects." Still others mentioned objective faults in the answers rather than faults in procedure. For example, one teacher mentioned that "minor details are included while leading points are omitted." Another stated that "applications are not as good as could be expected." Still another teacher simply stated that "the answers are inaccurate." General faults, such as "lack of practise in answering questions," "carelessness," "lack of concentration," "pupils do not know how to study," and so on were mentioned by one or more teachers.

It may be pointed out that teachers who fail to become definitely conscious of the difficulties which their students encounter will be unable to give them much definite constructive assistance in this phase of their learning. Probably the most significant conclusion to be drawn from the answers to this part of the questionnaire is that many teachers are failing to give attention to the procedure which students use in answering questions. They appear to be concerned largely with the accuracy of the answer and when it is wrong they fail to seek the cause in the procedure which the student has used.

Several of the faults which appear to be suggestive are given below. This list is not a complete statement of the faults of students in answering questions but should prove helpful to teachers who are desirous of ascertaining the reasons why students fail to answer questions satisfactorily.

Type 2.

The pupil fails to see real basis on which evaluation should be made.

Type 3.

A student seldom stays "on a single designated basis."

Type 3 or 4.

The pupil begins to answer the question before he has thought out what points should be included in the comparison. He often includes minor comparisons and omits the important points.

Unless there is a single designated basis for comparison, pupils fail to see all of the possibilities of comparison. The slothful pupil is content with one or two obvious comparisons.

The pupil can state some information about each thing compared, but omits the comparative connection between the two things. The pupil gives comments on one thing and omits the discussion on the other thing, showing how the two are alike comparatively.

In making comparisons, students usually give characteristics of things to be compared, or describe them separately but fail to make a complete comparison.

Type 4.

In comparing two things in general the worst fault is in the students' forgetting some of the important points to be compared. Often enough suggestions must be given so that the question finally should be placed under Type 3.

Pupils frequently fail to distinguish the important from the trival. The pupil's prejudices—likes and dislikes—influence the comparison.

Type 6.

Students confuse cause and effect. Method requires real *thinking and reasoning*—the hardest thing any student can be asked to do.

Students often know a statement is true but on account of lack of thoroness can not give causes or reasons.

Type 6 or 7.

Pupils, when asked, "Why?" seem to forget all reasoning and usually state the first thing that comes to mind and fail also to go far enough in explanation.

Type 7.

Pupils have difficulty in finding words to express the thought without repeating word for word parts of the passage to be explained. The chief difficulty seems to be the limitations of their vocabulary.

The trouble here is that pupils seem to have failed to study the statement under question sufficiently and so have quite an indefinite notion of the words used. And too, even if they succeed in getting a fairly good understanding of the text read, they are too careless of the choice of their own words in reproducing the meaning exactly. They are too hasty in the study of the passage and have not yet been trained to go slowly and get facts one at a time.

Type 8.

A discussion, like a summary often lacks point. The student fails to organize material and grasp essentials.

Students fail to properly organize data mentally before expressing them.

Type 9.

Here pupils are often prone to feel satisfied if only a few of the factors or qualities involved in the analysis are learned and reproduced. Many times but one factor will be offered when to completely analyze the problem will require several. Being allowed to "get off" with a minimum of thinking is one of the chief causes of poor analysis.

Type 10.

Pupils often fail to get the facts correlated before they attempt to state the existing relationships.

Pupils do not stop to analyze in order to see the relationship but answer without giving much thought.

Pupils answer various separate facts instead of showing relationship. These facts are usually true and have a bearing upon the question but the actual relationship is usually missing in the answer. That is, the pupil does not seem to see how a change in one fact would influence some other related fact.

Type 11.

Illustrations or examples. This is not so difficult after the student has had some practise, but at first he always tries to find his illustrations in some field remote from his own. It takes him a while to realize that scores of illustrations come daily under his observation—*if he observes.*

Type 12.

In classifying, the pupils make their answers too readily without first trying one class then another or without thinking whether the one they choose actually does fit better and why.

The student's failure in Type 12 is due mainly to his failure in Type 7. He fails to decide correctly upon a construction because he has not thought out carefully the exact meaning of the sentence. This in turn is due to mental laziness or carelessness.

Type 14.

He discusses each phase at random and separately without relating it to others.

A discussion like a summary often lacks point. The student fails to organize material and grasp essentials.

Students fail to properly organize data mentally before expressing them.

Students do not group their facts together sufficiently to discuss them properly.

Type 17.

A pupil doesn't really outline into main facts with heads under them, but makes long statements one after another, or puts steps in before their time.

General.

Pupils will not read and get a thoro knowledge of the subject-matter first.

With Types 8, 14, and 17 pupils do not summarize or discuss or outline in a way to make clear to another the things under discussion. They are not logical in the arrangement of their ideas. They answer, at least 85 percent of them, only well enough so that one who knows already can tell that they know, but only about 15 percent can arrange and subordinate well enough to be clear to the unintiated.

A common fault is careless reading of the question (or hearing it as the case may be). This is especially true in Algebra. Another fault is "saying the first thing that comes into mind" instead of carefully considering its application to the problem under discussion. In Geometry, for example, the student does not rapidly recall all theorems pertaining to the subject in hand, eliminate those inapplicable, and so discover the correct one.

Percent of unsatisfactory answers due to faulty procedure in answering questions. In Question V the teachers were requested to indicate separately for memory questions and for thought questions the percent of unsatisfactory answers which, in their opinion, were due primarily to bad habits of procedure. A summary of their replies is given in Table V. As previously indicated it appears certain that a considerable number of teachers have given relatively little thought to the faults in the procedure used by students in answering questions. Consequently, the replies to this question must be considered as representing no more than rough estimates. In the case of thought questions the median percent of unsatisfactory answers due to faulty procedure is approximately 50 for all of the subjects. In the case of memory questions faulty procedure is considered to be a less potent cause. There are two factors that probably tend to

TABLE V. PERCENT OF SATISFACTORY ANSWERS DUE TO FAULTY PROCEDURE IN ANSWERING QUESTIONS

(As Contrasted with Lack of Information)

PERCENT	THOUGHT					MEMORY				
	English	History	Science	Others	Total	English	History	Science	Others	Total
95	1	1			2	1	5			1
90	3	1	2		6	2		2		4
85			3	3	6	1			1	2
80	2	2		5	9	2				2
75	5	7	2	13	27	4	1	4	2	11
70			1	5	6			1	1	2
65	2		2		4					0
60	3	5	3	6	17	4	1		5	10
55					0					0
50	3	7	8	10	28	3	5	5	12	25
45	1				1					0
40		7	4	3	14	3	6	1	7	17
35	3	1	4		8		1	1	1	3
30	3	2	3	9	17				6	6
25	4	7		6	17	6	14	11	11	42
20		3	7	4	14	7	5	2	6	20
15					0	1	3	1	2	7
10	7				7	2	4	8	7	21
5	1	1		1	3	1	3	3	4	11
0				1	1					0
Median	48	48	48	51	50	40	25	25	25	26

make these estimates lower than they should be. Many teachers showed by their replies to Question IV that they had very little experience in analyzing answers for the purpose of determining the cause of faulty answers. Naturally they could not be expected to recognize the real extent of such faults of procedure on the part of the members of their classes. Other teachers use such a limited range in types of questions that they would not have a chance to observe as many faults in answer technique as teachers who use a greater variety of types of questions.

Relation between frequency of use and difficulty. There seems to be very little relation between the difficulty of a type and the frequency of its use. Type 6, causes and effects, is the most commonly used type and at the same time it, together with Type 13, application, is considered the most difficult for students. Only one of the five types reported as most difficult, Type 18, reorganization of facts, is among the five least frequently used; only one of the five types reported as least difficult, Type 13, comparison on a single basis, is among the five most frequently used.

Relation of frequency of use of types of questions to emphasis upon different types of learning. An earlier bulletin[1] of the Bureau of Educational Research reports the relative emphasis on different types of textbook study in the various subjects. We now have some data on the question, "How does the frequency of use of types of questions correspond with the emphasis on the different types of textbook study required of students?" Some of the types of study recognized in the report are too broad to be tested by any one type of thought question. There are, however, question types that correspond rather closely with certain of the study types.

To get a rough measure of the relation of the frequency of use of the different types of thought questions to the emphasis upon the types of learning, we may arrange the types in order of frequency of use and divide the list into quartiles (four equal parts). For instance, in English, Study Type II is in the first quartile (i. e. the three most frequently required study types); Study Type X is in the fourth

---

[1] Monroe, Walter S. "Types of learning required of pupils in the seventh and eighth grades and in the high school." University of Illinois Bulletin, Vol. 19, No. 15, Bureau of Educational Research Bulletin, No. 7, Urbana: University of Illinois, 1921, 16. p.

TYPES OF TEXTBOOK STUDY	TYPES OF QUESTIONS
II. Summary of central ideas in the lesson studied.	8. Summary.
III. Prepare an outline of principal points and supporting details in the lesson arranged to show order of relative importance and relationship.	17. Outline.
VI. Discovery of collateral or illustrative material for topics or problem under discussion.	11. Illustration or examples.
VIII. Appreciation of the significance of each word used in a concisely expressed statement or principle.	7. Explanation of the use or meaning of some phrase or statement in a passage.
IX. A clear comprehension of the essential conditions of a problem which is to be solved.	9. Analysis.
X. Discovery of new or supplementary problems related to the topic being studied.	19. Formulation of new questions.

quartile (i. e. one of the least frequently required study types).[1] Table VI shows the extent of agreement in rank in frequency of use of some corresponding types of study and questions. The first column gives by number the corresponding types as listed above. The first number given in the columns for the separate subjects designates the quartile rank of the study type on the same horizontal line; the second number designates the rank of the corresponding type of question. It will be noted that the rank agrees in seven out of eighteen possible cases. In English the disagreement is greatest in the case of questions calling for analysis (first rank) and study requirements of a clear comprehension of the conditions of a problem (fourth rank); in History, the discovery of collateral or illustrative material ranks second in study requirements but questions calling for illustrations or examples rank fourth in use; in Science, the same study requirement ranks fourth and the corresponding question type ranks first. Thirty-three percent of the Physics and Chemistry teachers in the first questionnaire mentioned Study Type X, discovery of new or supplementary problems, as a major type of study in these subjects, yet 44 percent of the teachers of these subjects by their admission do not ask any questions of Type 19, formulation of new questions, to their

---

[1] These facts are derived from Table II of the Bulletin referred to above.

TABLE VI. CORRESPONDENCE OF TYPES OF STUDY AND TYPES OF QUESTIONS

CORRESPONDING TYPES		ENGLISH		HISTORY		SCIENCE		AGREE	VARY	VARY WIDELY
Study	Quest.	Study Rank	Quest. Rank	Study Rank	Quest. Rank	Study Rank	Quest. Rank			
II	13	1	2	2	2	3	3	Hist.+Sci.	English	
III	17	2	3	1	2	4	4	Sci.	Eng.+Hist.	
VI	11	3	3	2	4	4	1	Eng.		Hist.+Sci.
VIII	7	2	1	3	3	3	2	Hist.	Eng.+Sci.	
IX	9	4	1	4	2	3	3	Sci.		Eng.+Hist.
X	19	4	4	3	4	2	3	Eng.	Hist.+Sci.	

high school juniors and seniors. It would seem from the data that some teachers may not be using the types of questions that would stimulate and test students' achievements in some of the types of study they are expected to use.

Conclusions. Probably the most significant conclusions to be drawn from this investigation are: (1) teachers are not sufficiently conscious of the types of questions which they are accustomed to ask and of the significance of these types, and (2) in general teachers do not analyze unsatisfactory answers to questions in order to ascertain whether such answers are due to a faulty technique on the part of the student.

A number of other conclusions are worthy of mention. (1) Teachers of the same subject vary widely in the extent of their use of different types of thought questions. (2) The frequency of the use of a given type does not seem to depend very much on (a) the school grade, (b) the subject, or (c) the supposed difficulty of the type. It is very likely that some teachers, who expect their students to use certain types of study, do not use the types of questions that are best suited to test their students' achievements due to those particular types of study. (3) Teachers individually think that certain types of questions are more difficult for students than others. However, there is no very great agreement among teachers as to the relative difficulty of the various types.

BULLETIN NO. 15

BUREAU OF EDUCATIONAL RESEARCH
COLLEGE OF EDUCATION

THE CONSTANT AND VARIABLE ERRORS OF EDUCATIONAL MEASUREMENTS

by

Walter S. Monroe
Director, Bureau of Educational Research

PRICE 25 CENTS

PUBLISHED BY THE UNIVERSITY OF ILLINOIS, URBANA
1923

UNIVERSITY
OF ILLINOIS
PRESS

TABLE OF CONTENTS

PAGE

Preface.. 5

Chapter I. Introduction.................................... 7

Chapter II. Causes, Nature and Magnitude of Constant Errors.. 10

1. Evidence of presence of constant errors................ 10
2. Constant errors in first trial scores.................... 18
3. Exact magnitude of constant error can not be determined 19

Chapter III. Causes, Nature and Magnitude of Variable Errors.. 21

1. Evidence of presence of variable errors................ 21
2. Method of describing variable errors.................. 21
3. Magnitude of variable errors to be expected in educational measurements................................. 24

Chapter IV. Effect of Constant and Variable Errors upon Derived Measures................................... 27

Chapter V. Effect of Errors upon Use of Educational Measurements.. 30

PREFACE

This bulletin, based upon a number of investigations conducted by the Bureau of Educational Research, brings together data relating to the errors encountered in educational measurements. Its purpose is to call the attention of users of educational tests to the nature and magnitude of the errors which they will encounter. The bulletin is not intended as a criticism of educational tests, but rather as an aid to a more intelligent use of them.

Walter S. Monroe, *Director.*

October 4, 1923

THE CONSTANT AND VARIABLE ERRORS OF EDUCATIONAL MEASUREMENTS

CHAPTER I

INTRODUCTION

Educational measurements for many generations have been made by means of written examinations and teachers' estimates. These, we have been told, are subject to very large errors. Standardized educational tests have been proposed as instruments for obtaining more accurate measures. These instruments, however, do not yield measures involving negligible errors. In our measurements of ability in reading, spelling, arithmetic and other school subjects, we have not and are not likely to approximate the accuracy and precision with which the scientist is able to measure height, volume, temperature, and mass. We obtain, in fact, errors very much greater than those with which we deal in ordinary physical measurements.

In order to avoid misleading interpretations of educational measurements, it is necesary for us to be familiar with the nature and significance of the errors which we encounter. We need also to have some concept of the absolute magnitude of these errors. In this bulletin we attempt to answer the following four questions with reference to the errors encountered in the measurement of achievement and general intelligence by means of standardized educational tests:

1. What are the causes that tend to produce the errors encountered in educational measurements?
2. What is the nature of these errors?
3. What is the magnitude of the errors to be expected?
4. What is the effect of these errors upon the average, standard deviation and coefficient of correlation?

Variations in testing conditions tend to produce errors in educational measurements. A pupil's performance and hence his score on an educational test depend upon a number of factors. For example, it has been found that recent instruction may operate to increase the scores of pupils. Impending school events or other distractions may tend to lower their scores. Among the factors which may be easily specified as influencing a pupil's score are the following: his emotional status, his physical condition, the effort which he makes, the set of his mind, the recency of instruction in the field of the test, the acquaintance which he has with the type of exercise he is asked to do, the manner in which the test is presented, the particular directions which are given him with reference to methods of work, the distractions, and the time allowed. There are other factors, such as the attitude of the teacher toward the test, which are more subtle in character but which doubtless in many cases operate to increase or decrease the scores of many or all of the pupils tested.

A test is standardized with reference to certain specific testing conditions. The use of a standardized educational test implies that these same testing conditions are to prevail when it is given to a group of pupils. This means that standard testing conditions must be secured for each pupil as well as for the group as a whole. If this is not done the norms do not constitute a valid basis for interpreting the scores. Any variations from the standard testing conditions tend to produce variations in the performances of some or all of the pupils. These variations constitute errors of measurement.

Constant and variable errors of measurement. Errors of measurement are of two types (1) constant errors and (2) variable errors.

A constant error is one which has the same magnitude for all of the scores of a given group. In other words the presence of a constant error results in all the scores of this group being either too high or too low. In the field of physical measurement we have an illustration of a constant error when a merchant gives short weights, such as a grocer who uses a peck or bushel measure which has a false bottom. The group of scores to which a given constant error applies may be those of a class, a school system or a group of school systems. It is possible that a given constant error would affect the scores made by boys and not those made by girls even when both sexes are tested together. Furthermore, it should be noted that a constant error may be either positive or negative.

A variable error of measurement is one which varies or differs in magnitude for the several scores of a given group. We may secure an illustration of variable errors in the field of physical measurement by having a group of persons guess the length of a given object, for example, a table or even a pencil. If these guesses are made independently they are found to extend over a considerable range. In order to determine the magnitude of the variable errors involved in any guess it is necessary to determine the true length. In our illustration this might be done by having the length carefully measured by means of a yardstick or a tapeline. However, if we have secured a reasonably large number of guesses we may obtain an approximately true measure of the length by taking the average of the guesses. The difference between the true measure and any guess constitutes a variable error.[1] Some of these differences are positive and some negative. A few approximate zero.

In the field of mental measurements it is generally not possible to obtain true scores. Hence, we cannot calculate the magnitude of the variable error in a given score, but the concept of the true score aids us in understanding the nature of the variable errors of measurement. Approximately half of the variable errors for a given group of scores are positive and approximately half negative. If they were assembled for a frequency distribution the shape would approximate the normal probability curve with the average at zero. For a few measures the variable error would be relatively large, either positive or negative, but most of them would be near zero.

Constant and variable errors of measurement occur simultaneously. The situation may be represented by the following equation:

Obtained score = true score + constant error + variable error.

In this equation both errors may be either positive or negative, or one positive and the other negative. However, a constant error will have the same sign for all members of a group, that is, if it is positive for one pupil it will be positive for all of the pupils. Variable errors change signs within the group. Altho the two errors occur simultaneously, it is helpful to consider them separately and to treat each independently of the other.

---

[1] This statement is not strictly accurate. As we shall point out in a later paragraph, constant errors and variable errors occur at the same time. Thus, the difference may be the algebraic sum of the constant error and the variable error. However, in our illustration from the field of physical measurements, it is unlikely that there will be a large constant error. In the field of mental measurements there may be a relatively large constant error.

CHAPTER II

CAUSES, NATURE, AND MAGNITUDE OF CONSTANT ERRORS

Evidence of the presence of constant errors in educational measurements. 1. Constant errors due to acquaintance with the test. It is obvious that if a test is new to a given group of pupils, one significant change in the testing conditions attends its second administration. The test is no longer new to the pupils even if a duplicate form is used. When it is given a third or fourth time there is an added acquaintance with the type of exercise and the general form of the test. The taking of a test in itself thus introduces a change in the testing conditions which can not be eliminated. In order to secure evidence of the constant error due to the effect of acquaintance with a test it is necessary only to give the test twice to the same group of pupils under testing conditions which otherwise are as nearly the same as possible and to compare the averages of the two sets of scores. The difference between the average of the first trial scores and the average of the second trial scores is an index of the magnitude of the constant error resulting from the change in the testing conditions due to the pupils' acquaintance with the test. This difference, however, should not be interpreted as being the true magnitude of the constant error of the second trial scores. It is possible that the first trial scores also involved a constant error due to failure to secure standard testing conditions. However, when the averages of the two sets of scores are not equal we have evidence of the presence of a constant error and an indication of its magnitude.[1]

The Illinois General Intelligence Scale[2] was given twice to several hundred pupils in Grades III to VIII inclusive. After making due allowance for the inequality of the two forms of this scale[3] the difference between the averages of the two sets of scores was approximately five points, or six months of mental age. In the eighth grade in which unusual testing conditions appear to have prevailed the difference was considerably greater. For Monroe's General Survey Scale in Arithmetic the difference between the average of the first trial scores and that of the second trial scores was approximately 3.2 points in Grades III to V, and 4.5 points in Grades VI to VIII. The writer recently had two forms of the Thorndike-McCall Reading Scale given to several groups of pupils. The average of the first trial scores (Form 2) was 47.78, the average of the second trial scores (Form 3), 51.69. Investigation has shown that these two forms are approximately equivalent. Hence, the difference between these two average scores indicates the magnitude of the constant error introduced by acquaintance with the form of the test.

---

[1] In case different forms of a test are used in the two applications, it will be necessary to inquire concerning their equivalence and to make an appropriate allowance for any lack of equivalence in comparing the two averages.

[2] Monroe, Walter S. "The Illinois Examination." University of Illinois Bulletin, Vol. 19, No. 9, Bureau of Educational Research Bulletin No. 6. Urbana: University of Illinois, 1921, p. 69.

[3] See page 10 of the bulletin just referred to.

One investigator[4] has reported data which is evidence of the presence of a constant error in the second trial scores yielded by the Burgess Picture Supplement Scale for Measuring Silent Reading Ability. Form 2 was given on the day following that on which Form 1 was used and care was exercised to secure as nearly the same testing conditions as possible. After discarding the records of all pupils who did not take both forms of the test the median scores for the two forms were 43 and 64. A part of the difference between these two median scores is undoubtedly due to the inequality of the two forms of the test used. This is shown by the fact that when the order of giving the two forms was reversed with another group of pupils, the median score dropped from 55 to 49. In a third group where Form 1 was given a few minutes after Form 2, the median score dropped from 58 to 49.

2. **Evidence of constant errors introduced by lack of equivalence of duplicate forms of a test.**[5] Altho the duplicate forms of a test are generally constructed so that they are expected to yield equivalent scores and to be used interchangeably, experience has shown that these forms are not always equivalent. Evidence of a

---

[4] Daley, H. C. "Equivalence of Forms 1 and 2 of the Burgess Picture Supplement Scale for Measuring Silent Reading Ability," Journal of Educational Research, 4:71, June, 1921.

[5] The lack of equivalence of duplicate forms of a test results in a constant error only when it is neglected as is the case when the same norms are used for interpreting the scores yielded by both forms or when comparisons are made between the scores yielded by the different forms without making due allowance.

TABLE I. DATA SHOWING THE EQUIVALENCE OF DUPLICATE FORMS OF TWO SILENT READING TESTS

Burgess Picture Supplement Scale			Thorndike-McCall Reading Scale		
Score	*Form 2*	*Form 3*	*Score*	*Form 2*	*Form 3*
20	1		66		6
19	1	1	63	3	9
18	6	1	60	5	9
17	7		57	17	19
16	7	4	54	14	37
15	4	5	51	46	18
14	9	4	48	36	65
13	12	9	45	99	72
12	18	18	42	81	61
11	20	15	39	43	60
10	27	21	36	44	49
9	30	23	33	27	20
8	26	33	30	6	4
7	17	31	27	6	2
6	19	25	24	1	
5	17	21	21	1	1
4	13	13			
3	9	19			
2	7	5			
1	7	6			
0	5	9			
Total.........	262	263		429	432
Median.......	9.37	8.08		45.18	45.78
Average.......	8.77	7.61		44.01	44.84

constant error due to such lack of equivalence is furnished by the illustration given in the preceding paragraph. More exact evidence may be secured by arranging the duplicate forms in alternate order, and distributing them in this manner to pupils as they happen to be seated in the classroom. Thus, if Form 1 and Form 2 are being compared the first, third and fifth pupils will have a copy of Form 1, the second, fourth, sixth, etc., of Form 2. If it is decided to secure information for three forms of one test at a time, a similar arrangement will result in every third pupil having a copy of the same form. Form 2 and Form 3 of the Thorndike-McCall Reading Scale were arranged in this way and given to several hundred children. The same procedure was followed with reference to the Burgess Picture Supplement Scale for Measuring Silent Reading Ability. The distribution of scores from the different forms of the two tests is given in Table I. Both the median and the average scores for the Thorndike-McCall Reading Scale show that Form 2 and Form 3 are approximately equivalent. In the case of the Burgess Picture

Supplement Scale, the differences between the medians and the averages indicate a lack of equivalence which can not be neglected safely when precise comparisons are being made between scores yielded by the two forms.

By a similar method the equivalence of the duplicate forms of the Illinois General Intelligence Scale, Monroe's General Survey Scale in Arithmetic and Monroe's Standardized Silent Reading Tests, Revised, was studied.[6] The evidence collected for these three measuring instruments indicates that the different forms are slightly lacking in equivalence. This is especially true of the measures of rate yielded by the silent reading test. Thus, it has been considered necessary to give correction numbers whereby the scores yielded by one form of the test may be reduced to a basis comparable with those yielded by the other forms.

A similar study of the three forms of Monroe's Standardized Silent Reading Tests indicated a marked lack of equivalence.[7] In order to eliminate the constant error due to this cause corrections have been calculated which may be used to reduce the scores yielded by the different forms to a comparable basis. Separate sets of norms have been stated for each form.

3. **Evidence of constant errors due to instruction functioning as coaching.** When any considerable period of time elapses between two trials on a given test, the instruction which pupils receive during this interim may materially influence their second trial scores. In a recent investigation[8] by the Bureau of Educational Research it was found that for a group of 134 children the increase of the second trial scores on the Illinois General Intelligence Scale over the first trial scores was equivalent to slightly more than four years in mental age. The two trials were six months apart and hence the normal increase to be expected would be six months. If we assume that the first trial scores were accurate, it follows that the constant error introduced in the second trial scores was in the neighborhood of three

---

[6] Monroe, Walter S. "The Illinois Examination." University of Illinois Bulletin, Vol. 19, No. 9, Bureau of Educational Research Bulletin, No. 6. Urbana: University of Illinois, 1921, 70 p.

[7] Monroe, Walter S. "Report of Division of Educational Tests for '19-20." University of Illinois Bulletin, Vol. 18, No. 21, Bureau of Educational Research Bulletin No. 5. Urbana: University of Illinois, 1921, p. 19.

[8] Odell, Charles W. "The use of intelligence tests as a basis of school organization and instruction." University of Illinois Bulletin Vol. 20, No. 17, Bureau of Educational Research Bulletin, No. 12. Urbana: University of Illinois, 1922. 78 p.

and one-half years of mental age. Investigation revealed that the teachers of these pupils had given instruction which incidentally functioned as coaching and increased the scores on the second trial of the test.

In an unpublished study made by Mr. H. N. Glick, deliberate coaching on the Army Group Intelligence Scale Alpha was given to a number of pupils. The increases in the scores when the test was repeated amounted in some cases to several hundred percent of the original scores.

In dealing with measures of achievement it is more difficult to demonstrate the presence of constant errors. Instruction is expected to result in an increase in achievement. However, the use of a standardized educational test implies the existence of standard conditions with respect to recency of instruction. Furthermore, in many cases we are measuring merely a sample of a pupil's achievement. In such cases our measurements are valid only if this sample is representative. Hence the increase in the scores yielded by a second application of an achievement test may represent a combination of true growth and spurious growth. If the instruction has been very recent in the field of the test or if it has been concentrated upon the particular sample of achievement which the test measures directly the increase in scores will represent, for the most part, spurious growth. Additional evidence on this point will be given in the next section.

When two dimensions of a pupil's ability are measured separately as in the case of both rate and comprehension of silent reading, we find that frequently the magnitude of one dimension is increased at the expense of the other. This may be due to the instruction which pupils have received or to other directions given them at the time of taking the test. Unless the two dimensions are interpreted together the effect of such compensating relation will be similar to that of a constant error.

4. **Evidence of constant errors in measures of progress in educational experimentation.** When we attempt to secure a measure of progress in achievement in a school subject by taking the difference between the averages (or medians) of two sets of scores, we frequently find evidence that one or both sets of scores involves an unknown constant error. Table II gives certain gains in achievement which were obtained in an experiment to determine the relative effect

TABLE II. TWO SETS OF GAINS IN ACHIEVEMENT WHICH INDICATE THE PRESENCE OF CONSTANT ERRORS IN CERTAIN SETS OF SCORES, FIFTH GRADE

Group	No. of Pupils	Reading Rate		Reading Comprehension		Arithmetic	
		I	II	I	II	I	II
I	70	27.93	−15.78	.96	.35	23.82	21.45
II	72	3.67	22.11	1.21	1.86	14.72	5.44
III	326	− 4.77	33.25	.92	2.06	12.07	6.36
IV	133	− 6.60	22.90	.82	.95	17.04	10.09
V	157	9.29	27.35	1.48	2.12	10.65	5.83
VI	143	− 9.26	41.48	.08	2.36	4.69	5.38

of the number of sections into which a class was divided.[9] The six experimental groups were taught under the same conditions with the one exception regarding the number of sections into which the classes were divided. The tests used were Monroe's Standardized Silent Reading Test I, Revised, and Monroe's General Survey Scale in Arithmetic. Form 1 of these tests was given early in October, Form 2, the first of February, and Form 1 was repeated early the following May. The first gains were calculated by subtracting the average of the October scores from that of the February scores; the second, by subtracting the average of the February scores from that of the May scores. The two forms of these tests have been shown to be slightly lacking in equivalence, especially in the case of reading rate.[10] The gains in Table II, however, are evidence of the presence of constant errors in addition to those resulting from the slight non-equivalence of the different forms.

On the basis of our knowledge of the effect of practise we should expect the first gains to be larger than the second gains unless the variations of experimental conditions materially influenced the achievements of the pupils, which is extremely unlikely. We find in both reading rate and reading comprehension that the first gains are frequently less than the second. In three cases the first gain for reading rate is negative. In arithmetic the first gains are larger than the second in all cases except one. The smaller gains during the first semester than the second and particularly the negative gains are evidence of the presence of a constant error in at least one of the sets of scores from which the gains were computed. The gain in reading rate shown for Group I is interesting; from October to February there is a very marked increase in rate; for the second semester the gain is negative. This suggests that the average February score was too large, i.e., it involved a positive constant error.

---

[9] Monroe, Walter S. "Relation of sectioning a class to the effectiveness of instruction." University of Illinois Bulletin, Vol. 20, No. 11. Bureau of Educational Research Bulletin, No. 11. Urbana: University of Illinois, 1922. 18p.

[10] Monroe, Walter S. "The Illinois Examination," University of Illinois Bulletin, Vol. 19, No. 9, Bureau of Educational Research Bulletin No. 6. Urbana: University of Illinois, 1921, p. 12-18.

Similar evidence of the presence of constant errors in measures of achievement is found in a recent study of the relation of class size to school efficiency.[11] In this investigation, as in the one just described, experimental groups were arranged in pairs with the experimental conditions alternating in the two semesters. Especially in Grades V and VII the relative magnitude of gains made in the different semesters indicates the presence of a constant error in at least one set of scores from which the gains were computed.

In another investigation[12] conducted by the Bureau of Educational Research the average increases in mental age during a period of six months for two groups of children, each numbering about 3000, were found to be .4 years and .9 years. During the next six months for the same two groups the increases were 1.4 years and 1.0 years respectively. The normal increase in mental age during either of these intervals is of course six months. The obtained increase for the first period might be expected to be somewhat greater because of the presence of a constant error introduced by the general practise effect. However, in one case the difference between the first and second trial scores is less than six months and in both the increase is less than the corresponding differences between the second and third trial scores. No explanation was found for these inconsistent gains but they are evidence that in some way an unknown constant error was introduced in some if not in all of the scores. The facts of this illustration become even more striking when we note that the total of the two gains for the first group is 1.8 years and that for the second 1.9. Thus, when the total interval of twelve months is considered, the total increase in mental age is approximately the same for the two groups. On the other hand, if the two intervals of six months are taken, the increases in the mental age are radically different for the two groups.

---

[11] "Relation of size of class to school efficiency." University of Illinois Bulletin, Vol. 19, No. 45, Bureau of Educational Research Bulletin No. 10. Urbana: University of Illinois, 1922, p. 20.

[12] Odell, Charles W. "The use of intelligence tests as a basis of school organization and instruction." University of Illinois Bulletin, Vol. 20, No. 17, Bureau of Educational Research Bulletin No. 12. Urbana: University of Illinois, 1922. 78 p.

In the same investigation if only the scores yielded by the Illinois General Intelligence Scale are considered, the gain between the first and second testings is 1.1 years. For the second period it is 1.4 years. A constant error due to practise effect is expected in these gains but it is surprising to find that the second gain, which is the difference between the second and third trial scores, involves the larger error.

In each of these illustrations we have evidence of the presence of a constant error for which the cause is obscure. Furthermore, the exact magnitude of the constant error is unknown. The obscurity of the cause is due in part to the large number of teachers and pupils participating in each of these educational experiments. The constant errors may have been due to changes in the interest and attitude of the teachers and pupils toward the test. However, it was not possible to secure any direct evidence on this point. The fact that the cause is obscure makes the possible presence of constant errors in such data a serious matter and tends to arouse suspicions regarding the accuracy of measurements of ability in large cooperative experiments.

5. **Evidence of constant errors in subjective scoring.** The evidence cited in the preceding pages has related to testing conditions. The scoring of the tests used was highly objective. In case the scoring of the test papers is not objective it is necessary to consider also the constant errors which may be introduced in this process. In the marking of examination papers and other pupil performances where the scorer is asked to exercise judgment, much evidence has been collected to show that two persons differ widely in the scores which they assign to the same pupil performances. These differences are due in part to the presence of a constant error resulting from the fact that one of the scorers tends to be more liberal than the other. In a recent investigation[13] several sets of pupil performances for which the scoring was rather highly subjective, were scored inde-

---

[13] Monroe, Walter S. "A critical study of certain silent reading tests." University of Illinois Bulletin, Vol. 19, No. 22. Bureau of Educational Research Bulletin No. 8. Urbana: University of Illinois, 1922. 52 p.

TABLE III. SUBJECTIVITY OF SCORING REPRODUCTIONS BY THE WORD-COUNTING METHOD

Test	Form	Grade	No. of scores	Scorers	Difference of average scores
Memory................	I	IV	92	Y—C	− 9.9
Memory................	I	IV	27	Y—K	− 5.1
Memory................	II	IV	116	Y—C	− 2.0
Memory................	I	VII	123	Y—K	− 7.5
Memory................	II	VII	100	Y—C	− 8.2
Memory................	II	VII	31	Y—K	+ 4.1
Reproduction...........	I	IV	94	L—K	+ 6.8
Reproduction...........	II	IV	31	L—C	— 1.6
Reproduction...........	II	IV	68	L—K	+ 4.7
Reproduction...........	I	VII	117	M—F	− 0.5
Reproduction...........	II	VII	113	F—C	− 6.0
Brown..................	I	IV	111	T—My	+12.8
Brown..................	II	IV	110	T—My	+ 6.9
Starch (No. 7)...........	I	VII	119	M—C	− 5.8
Starch (No. 6)...........	II	VII	121	M—C	− 2.0

pendently by two persons under the supervision of a third. A part of one table is reproduced from this report to furnish evidence of the presence of a constant error in the scores assigned by one or both of the scorers. The entries in the column headed "difference of average scores" were obtained by subtracting the average of the scores assigned by the second scorer from the average of those assigned by the first scorer. Some of these differences are relatively large. It appears that the scorer is not always consistent with respect to his constant error. Scorers Y and K show positive differences for one set of papers and negative differences for another set.

In the same investigation, eighty-six compositions were rated independently by two persons using the Willing Scale for Measuring Written Composition. The difference between the averages of their scores was 6.7.

Constant errors in first trial scores. As we have already indicated, first trial scores may involve constant errors. If there have been any departures from standard testing conditions we may expect to find the scores yielded too high or too low. It is possible to coach pupils for the first administration of a test as well as for later administrations. It may happen that where there has been no intentional coaching the instruction which they have received immediately prior to the taking of the test has served as preparation for the test.

Furthermore, if the norms are for pupils who are relatively unacquainted with testing procedure, test scores made by pupils who are accustomed to taking tests will involve a constant error with reference to these norms. At first the norms for our standardized educational tests were based upon scores obtained from pupils who had little or no experience in the taking of tests. This was necessarily so because such tests were new. As tests have become more widely used this factor of the testing conditions has changed, and it is probably true that norms for tests which have been recently standardized are based upon scores from many pupils who are familiar with general testing procedure. However, we have no specifications concerning the degree of acquaintance with the testing procedure for which the norms are stated.

In addition to the influence of instruction and acquaintance with testing procedures, constant errors may be introduced in first trial scores by the attitude of the pupils toward the test, by the way in which the test is explained to the pupils, and by a number of other factors which are subject to only partial control. In the case of handwriting the performances of pupils are very easily influenced by the type of directions given them. For example, in response to the instructions "Write as fast as you can" one college sophomore increased her rate of writing 77 letters per minute over her rate when writing for highest quality. One investigator[14] has presented evidence which shows that if pupils know they are being tested they will tend to write much more slowly than their normal rate. This reduction in rate is usually accompanied by an increase in quality. Similar results have been found for tests in other fields. The fact that test scores are influenced in this way by the directions given to pupils does not mean that they necessarily involve a constant error. It is only when these directions constitute departures from the standard testing conditions that we may expect constant errors. The evidence presented here merely shows what may happen when there are even slight departures from standard testing conditions.

Exact magnitude of constant errors can not be determined. In none of the cases cited to illustrate the presence of constant errors, was it possible to determine the exact magnitude of the constant error unless some basis for comparison was assumed. When a test is repeated after a short interval of time the difference between the averages of the scores obtained from the two trials becomes the magnitude of the constant error in the second trial scores only if the first trial scores involve no constant error. Such an assumption may be justified in certain cases but one can never be certain that standard testing conditions prevailed in all details. Even when the examiner has exercised special care some of the more subtle factors of the testing conditions may not have been completely controlled. Unless good evidence can be produced in support of the assumption that the first trial scores involved a negligible constant error it is not safe to consider the difference between the averages of the two sets of scores as equivalent to the constant error. In more complex situations where a test is given three or more times for the purpose of measuring progress for two or more periods, it becomes more obvious that the exact magnitude of the constant error can not be determined. This condition has been indicated already in the evidence presented to show that constant errors were introduced in the data gathered in large cooperative educational experiments.

---

[14] Sackett, L. W. "Comparable measures of handwriting." School and Society, 4:640-45, October 21, 1916.

Altho one can not determine the exact magnitude of the constant error of measurement in a given case he can frequently present evidence to show that it probably does not exceed a certain amount. If his use of the data does not involve precise comparisons it may be possible to show that the constant error may be safely neglected. However, when precise comparisons are required and conclusions depend upon small differences between average or median scores the possible presence of constant errors makes such conclusions of doubtful validity.[15]

---

[15] In order to contrast the effects of the two types of errors upon the average and other derived measures, the consideration of the fourth question stated on page 7 with reference to constant errors is left until after the treatment of variable errors. The effect of both types of errors upon derived measures is considered in Chapter IV.

CHAPTER III

CAUSES, NATURE, AND MAGNITUDE OF VARIABLE ERRORS OF MEASUREMENT

1. **Evidence of variable errors of measurement secured when a test is repeated.** In order to secure evidence of the presence of variable errors of measurement it is necessary only to repeat a test after a short interval of time and compare the two scores of individual pupils. When this is done it is found that some pupils make a higher score on the first test and others on the second. In Table IV, two sets of scores yielded by the Monroe General Survey Scale in Arithmetic are given. The first pupil made a score of 51 on the first trial and 59 on the second. The difference in the two scores is — 8. Most of the differences are small. A few are relatively large. Approximately half are positive. The facts shown in this table are typical of the scores yielded by educational tests. For a few tests the scores involve smaller variable errors of measurement but for a number they are larger than in this illustration.

It should be noted that the differences between the two scores given in Table IV are not the variable errors of measurement. They are merely indicative of the presence of such errors and, in a crude way, of their magnitude. Neither set of scores can be considered true scores. Both are subject to variable errors and also, possibly, to an unknown constant error. In order to obtain the exact magnitude of the variable error of measurement for a given pupil it would be necessary to secure a true score and to subtract the obtained score from it. Such differences, when corrected for the constant error, would be the variable errors of measurement.

Method of describing the variable errors of measurement. As we have already indicated, it is impossible to determine the pupil's true score (see page 9). Furthermore, we always find variable errors and constant errors occuring in combination. It is, however, possible to secure a description of the magnitude of the variable errors of measurement which may be expected in the scores yielded by a given educational test when it is administered under standard testing conditions. Two sets of scores such as given in Table IV furnish the data upon which this description is based. These are obtained by

TABLE IV. SCORES YIELDED BY TWO APPLICATIONS OF MONROE'S GENERAL SURVEY SCALE IN ARITHMETIC TO A GROUP OF FIFTH GRADE PUPILS

Form I	Form II	Difference	Form I	Form II	Difference
51	59	−8	46	58	−12
49	60	−11	49	54	−5
46	60	−14	60	71	−11
77	84	−7	43	41	+2
42	43	−1	45	40	+5
43	43	0	30	24	+6
51	63	−12	28	23	+5
33	36	−3	46	38	+8
41	48	−7	34	32	+2
40	46	−6	59	56	+3
39	53	−14	63	72	−9
35	47	−12	42	45	−3
42	47	−5	45	53	−8
21	25	−4	48	73	−25
113	109	+4	51	59	−8
42	35	+7	53	56	−3
11	8	+3	68	68	0
23	21	+2	24	19	+5
45	38	+7	45	66	−21
54	49	+5	21	21	0
21	31	−10	37	36	+1
53	48	+5	38	39	−1
43	53	−10	39	50	−11
106	86	+20	50	64	−14
27	19	+8	30	43	−13
46	45	+1	33	42	−9
42	29	+13	65	74	−9
45	62	−17	69	86	−17
55	54	+1	51	59	−8
46	35	+11	43	43	0
38	32	+6	53	52	+1
17	15	+2	38	35	+3
53	61	−8	62	72	−10
52	65	−13	77	75	+2
50	58	−8	85	76	+9
41	43	−2	111	95	+16
37	48	−11	70	66	+4
26	34	−8	107	84	+23
57	65	−8	9	8	+1
51	59	−8	27	25	+2
34	41	−7	39	32	+7
26	36	−10	25	23	+2
22	24	−2	59	55	+4
59	64	−5	57	47	+10
49	37	+12	41	36	+5
61	75	−14	56	62	−6

two applications of the same test or of duplicate forms of a test to a group of representative pupils. These two applications should be separated by only a small interval of time. One type of description of the magnitude of the variable error of measurement is obtained by calculating the coefficient of correlation between these two sets of scores. This, when applied to an educational test, is called the coefficient of reliability. It indicates in a rough way the magnitude of the variable error of measurement, and is unaffected by the presence of any constant error of measurement in the two sets of scores from which it is calculated.

The coefficient of reliability an unsatisfactory description of variable errors of measurement. Altho the coefficient of reliability is an index of the variable error of measurement, a given coefficient, say .85, can not be interpreted directly in terms of the magnitude of these variable errors of measurement. Experienced persons are able to attach a reasonably concrete meaning to a given reliability coefficient but to an inexperienced person a coefficient of reliability, say .72 or .95, can have little more than a very general meaning.

Under certain conditions reliability coefficients furnish us with an index of the relative magnitude of the variable errors to be expected in the scores yielded by different tests. For example, if the reliability coefficient for one test is .65 and for another, .85, we should expect to find the variable errors of measurement for the second test much smaller. However, considerable caution must be exercised in comparing coefficients of reliability. The writer has shown[1] that the correlation of the two scores yielded by a given test is much smaller when the scores are taken from a single grade than when taken from a sequence of two or more grades. In one illustration when the scores were assembled separately for half-grade groups the highest coefficient of correlation was .57. For one half grade it was .12 and for another .27. When the scores for all grades from III-B to VIII-A, inclusive, were assembled the coefficient of correlation was .76. In this illustration there were from two hundred to four hundred cases in each half grade. Hence the variations can not be explained on the basis of sampling.

Probable error of measurement used to describe the variable error. Altho the coefficient of reliability is unsatisfactory it may be used as a basis for calculating the probable error of measurement. The formula[2] for this is given below.

---

[1] Monroe, Walter S. An Introduction to the Theory of Educational Measurements. Boston: Houghton Mifflin Company, 1923, p. 356.

$$P.E._M = .6745 \frac{\sigma_1 + \sigma_2}{2} \sqrt{1 - r_{12}}$$

In this formula r_{12} is the coefficient of correlation between first and second trial scores. σ_1 is the standard deviation of the distribution of the first trial scores and σ_2 a corresponding measure for the second trial scores.

It should be noted that the probable error of measurement does not give the magnitude of the variable error of measurement for any one pupil. It gives merely the limits between which we may expect to find 50 percent of the variable errors of measurement of a given group of scores. For example, the probable error of measurement for the rate score yielded by the Courtis Silent Reading Test No. 2 has been found to be 19.3.[3] This means that 50 percent of the variable errors of measurement were greater than 19.3, approximately one-half of them being positive. It also means that 50 percent of them were not larger than 19.3 nor smaller than — 19.3. In the case of a given pupil we can state only the chances that the probable variable error of measurement of his score does not exceed certain limits; as, for example, in the Courtis Silent Reading Test referred to, the chances were just even that the variable error of measurement in his score was not larger than ± 19.3. The chances are 4.6 to 1 that the variable error of measurement of his score is between — 38.6 and +38.6. The chances for other limits also may be stated.

The magnitude of variable errors to be expected in educational measurements. In the data given to show the presence of both constant and variable errors in our educational measurements their magnitude has been indicated. It is clear that they are much greater than the corresponding errors in physical measurements. In another place the writer has discussed the relative magnitude of the errors in the scores yielded by standardized tests and the errors in the grades assigned to examination papers.[4] The evidence presented indicates that the variable errors of measurement for a number of widely used standardized educational tests are only slightly less than the variable errors of measurement for written examinations. Some additional data with reference to the variable errors of the scores yielded by standardized tests will be helpful in arriving at the true understanding of their magnitude.

---

[2] For an explanation of this formula see, Monroe, Walter S. An Introduction to the Theory of Educational Measurements. Boston: Houghton Mifflin Company, 1923, p. 347-56.

[3] Monroe, Walter S. "A critical study of certain silent reading tests." University of Illinois Bulletin, Vol. 19, No. 22, Bureau of Educational Research Bulletin No. 8. Urbana: University of Illinois, 1922, p. 33.

In a critical study of a group of silent reading tests[5] it was shown that the probable error of measurement for some tests was greater than 25 percent of the average score. In fact, for Brown's Silent Reading Test it was found to be more than 50 percent. In the tests which make up the Illinois Examination, only twelve of forty-two probable errors of measurement which were calculated were greater than 10 percent of the average score.[6] The authors of the Stanford Achievement Test[7] announce that the probable error of measurement for this battery of tests is approximately two months of educational achievement. The coefficients of reliability are high. It is likely that these authors have succeeded in reducing the variable errors of measurement to a lower minimum than has been secured by others. This has been accomplished in part through extending the length of the test.

Using scores which were the medians of eight independent ratings of English compositions by means of the Nassau County Supplement to the Hillegas Scale, Hudelson[8] has given coefficients of reliability ranging from .69 to .84. The writer has estimated that if the ratings of a single judge had been used the coefficients of correlation would have been in the neighborhood of .40 instead of ranging from .69 to .84, and the probable error of measurement would have been a little more than six-tenths of one step of the scale used. This may appear relatively small but when we examine the norms we find that the unit used is relatively large. The average increase in norms from the fourth to twelfth grades, inclusive, is only slightly more than four-tenths of a unit per year. Between the eighth and ninth grades the increase is only two-tenths of a unit. The greatest yearly increase is six-tenths of a unit. Thus we have here a probable variable error of measurement which is relatively large.

---

[4] Monroe, Walter S., and Souders, L. B. "The present status of written examinations." University of Illinois Bulletin, Vol. XXI, No. 13. Bureau of Educational Research Bulletin No. 17. Urbana: University of Illinois.

[5] Monroe, Walter S. "A critical study of certain silent reading tests." University of Illinois Bulletin, Vol. 19, No. 22, Bureau of Educational Research Bulletin No. 8. Urbana: University of Illinois, 1922, p. 52.

[6] Monroe, Walter S. "The Illinois Examination." University of Illinois Bulletin, Vol. 19, No. 9, Bureau of Educational Research Bulletin No. 6. Urbana: University of Illinois, 1921, p. 49.

[7] Kelley, T. L., Ruch, G. M., and Terman, L. M. Stanford Achievement Test Manual of Directions. Yonkers: World Book Company, 1923.

[8] Hudelson, Earl. "English composition, its aims, methods, and measurement." Twenty-second Yearbook of the National Society for the Study of Education, Part I. Bloomington, Illinois: Public School Publishing Company, 1923, p. 62.

CHAPTER IV

THE EFFECT OF CONSTANT AND VARIABLE ERRORS UPON DERIVED MEASURES

The effect of errors of measurement upon derived measures. There seems to be a prevailing idea that the effect of errors of measurement upon such derived measures as the average, median, standard deviation and coefficient of correlation, may be safely neglected if the derived measure is based upon a sufficiently large number of cases. This is only partially true. A constant error in the original data makes the average in error by the amount of the constant error. Any increase in the number of cases has no effect upon the magnitude of the constant error. It can not be eliminated or even reduced unless we are able to determine its magnitude, in which case it may be subtracted from the average. The same situation prevails for the median. However, a constant error does not affect the standard deviation and other measures of variability. Neither does it affect the coefficient of correlation.

As we have already pointed out in the preceding pages, we are seldom able to determine even approximately the magnitude of the constant errors of the scores yielded by educational tests. We have evidence only of their presence. Hence, it is impossible to make any accurate correction for a constant error. It has been estimated that second trial scores are about 10 percent larger than first trial scores but studies of different tests indicate that this constant error of measurement varies widely. For some tests the difference between first trial scores and second trial scores is much less than for others. It is also doubtless less for some groups of pupils than for others. When pupils are acquainted with the testing procedure the increase of second trial scores over first trial scores may be very slight, especially if the children are acquainted with other tests having similar structure. When compared with first trial scores, third trial scores involve a somewhat larger constant error than those obtained from the second trial, and beyond the third trial it is likely that there is some increase.

Unlike constant errors of measurement variable errors tend to neutralize each other in the average. The reason for this is easily understood because approximately one-half of the variable errors of measurement are negative and the other half positive. If we increase the number of cases the magnitude of the variable errors of measurement in the average is decreased. The relation is given by the following formula:

$$\text{P.E.}_{\text{M}} \text{ average} = \frac{\text{P.E.}_{\text{M}}}{\sqrt{N}}$$

It should be noted that the error of the average due to the presence of variable errors of measurement in the data can not be explicitly defined. It is necessary to describe it in terms of the probable error (P.E._{M} average). The above formula gives the limits between which the chances are even that the error of the average will fall.

Variable errors of measurement tend to make the standard deviation and other measures of variability larger than they would be otherwise. The relation between the obtained standard deviation and the true standard deviation is given by the following formula:

$$\sigma_{\text{true}} = \sigma_{\text{obtained}} \sqrt{r_{12}}$$

In this formula r_{12} is the coefficient of reliability of the scores concerned. Since N does not appear in this formula it follows that increasing the number of cases does not have any effect upon the variable error of measurement of a measure of variability.

The presence of variable errors of measurement in our data always tends to decrease the coefficient of correlation.[1] If each of the two sets of facts whose relationship we are studying has been measured in duplicate, it is possible to correct for the effect of these variable errors of measurement. For example, if it is desired to secure the true correlation between ability to reproduce a selection read and ability to answer questions upon it, we may secure a corrected coefficient of correlation by measuring each of these abilities twice. One formula which has been used for this purpose is the following:

$$r_{pq} = \frac{\sqrt{(r_{p_1q_2})\ (r_{p_2q_1})}}{\sqrt{(r_{p_1p_2})\ (r_{q_1q_2})}}$$

r_{pq} here indicates the true correlation between two series of measures, p and q, of the facts A and B.

p_1 and p_2 are two independent measures of A.

q_1 and q_2 are two independent measures of B.

---

[1] Thorndike, E. L. Theory of Social Measurements. New York: Teachers College, Columbia University, 1916, second edition, p. 178.

$r_{p_1q_2}$ is the correlation obtained from the first measure of A and the second measure of B.

$r_{p_2q_1}$ is the correlation obtained from the second measure of A and the first measure of B.

$r_{p_1p_2}$ is the correlation between the two measures of A.

$r_{q_1q_2}$ is the correlation between the two measures of B.

In a recent study[2] it was desired to secure the correlation between the comprehension in silent reading as measured by Monroe's Standardized Silent Reading Tests with the scores yielded by a test of memory. It was recognized that both sets of scores involved variable errors of measurement which would materially decrease the magnitude of the coefficient of correlation. For this reason it was arranged to measure each trait in duplicate. The coefficients of correlation obtained from correlating each measure of comprehension yielded by a silent reading test with the two measures of memory were .31, .33, .31, and .35. The coefficient of reliability for the memory test was .35 and for the silent reading test, .73. In this illustration $r_{p_1p_2}$ equals .65, $r_{q_1q_2}$ equals .35, $r_{p_1q_2}$ equals .33, and $r_{p_2q_1}$ equals .31. Substituting these values in the formula given above we obtain for the corrected coefficient of correlation between comprehension and memory a value of .67. This gives some indication of the effect of the variable errors of measurement in these two sets of scores upon the coefficient of correlation between comprehension and memory.

---

[2] Monroe, Walter S. "A critical study of certain silent reading tests." University of Illinois Bulletin, Vol. 19, No. 22, Bureau of Educational Research Bulletin, No. 8. Urbana: University of Illinois, 1922, p. 41.

CHAPTER V

EFFECT OF ERRORS UPON USE OF EDUCATIONAL MEASUREMENTS

The attitude toward educational measurements as affected by the recognition of errors. The effective use of any instrument depends upon a frank recognition of its limitations. Standardized educational tests are no exception. They have been advertised by their authors and by others as measuring instruments which are distinctly superior to ordinary written examinations. At first attention was centered upon their use and few studies were made of the errors involved in the scores yielded. It has been apparent for some time, however, that test scores are subject to errors which sometimes are astonishingly large. This bulletin has been written to call attention to the nature of these errors, their magnitude, and their effect upon the average and other derived measures.

Those of us who have not been concerned with the errors involved in educational measurements may tend to feel doubtful of the value of standardized tests after realizing the nature and magnitude of the errors encountered. In the judgment of the writer, this effect should not be produced. We cannot expect to make our use of educational tests most effective until we are informed concerning the limitations of the measures yielded. A frank recognition of the presence of both constant and variable errors should enable the users of educational tests to do their work more efficiently. In many cases they can avoid erroneous interpretations which they otherwise would make. Furthermore, it is only by understanding the nature of the errors which are likely to be encountered that we can take steps to reduce them to the lowest possible minimum, and be aided in the construction of improved measuring instruments because of our knowledge of the defects of the present ones.

The writer does not advise the discontinuance of the use of standardized educational tests because the measures yielded have been shown to involve both variable and constant errors larger than many of us supposed. There is abundant evidence to show that the use of educational tests in our schools is increasing their efficiency. Still greater improvement may be expected when measuring instruments are used more intelligently.

BULLETIN NO. 16

BUREAU OF EDUCATIONAL RESEARCH
COLLEGE OF EDUCATION

AN ANNOTATED BIBLIOGRAPHY DEALING WITH THE CLASSIFICATION AND INSTRUCTION OF PUPILS TO PROVIDE FOR INDIVIDUAL DIFFERENCES

by

CHARLES W. ODELL
Associate, Bureau of Educational Research

PRICE 50 CENTS

PUBLISHED BY THE UNIVERSITY OF ILLINOIS, URBANA
1923

UNIVERSITY
of ILLINOIS
PRESS

PREFACE

The making of adequate provisions for individual differences represents one of the most important problems which educators are now facing. The need is beginning to be generally recognized and numerous procedures are being tried out in our schools. Many reports of attempts to provide for individual differences are being made. In order that one who is interested in making such provisions may profit by the experience of others it is necessary that he acquaint himself with the work which they have done. In order to assist the practical school man in this endeavor Dr. C. W. Odell has brought together an extended bibliography on "The Classification and Instruction of Pupils to Provide for Individual Differences." To each of the references he has added a brief annotation which indicates the character of the information which will be found in the reference. Altho the list probably is not complete it is believed to be more so than any other bibliography on this topic which is now available.

Walter S. Monroe, *Director*.

November 10, 1923

An Annotated Bibliography Dealing with the Classification and Instruction of Pupils to Provide for Individual Differences

The bibliography which follows was prepared by the writer in connection with his work described in a previous bulletin of the Bureau of Educational Research entitled "The Use of Intelligence Tests as a Basis of School Organization and Instruction," and may be considered in the light of an appendix to that bulletin. It has, however, been revised and brought up to date, that is practically to the end of the school year of 1922-23. No claim is made for its absolute completeness nor that the references given are more important than those not included. An effort has been made to list all important and most unimportant references that have appeared in the twenty-five or thirty leading educational periodicals of this country. In addition to these references many others are included from books, school reports, bulletins, proceedings of various educational bodies, etc. The writer did not make a thoro search through all of the reports of city superintendents and other school officials, proceedings of state and other teachers' associations and various other similar sources.

The references given have been classified into two main divisions. The first part of the bibliography deals with discussions of and provisions for individual differences that involve the use of standardized tests of general intelligence or achievement. The second part contains references about provisions and discussions that do not involve the use of such tests. Articles that deal with both sorts of classification, that based on the results of tests and that upon some other basis, are included in the first part.

PART I.

1. Adler, Martha. "Mental tests used as a basis for the classification of school children," Journal of Educational Psychology, 5:22-28, January, 1914.

In a New York school 160 first and fourth-grade pupils were tested with the Goddard Revision. Almost 50 percent of those tested were placed in advanced sections and of these about 60 percent gained one-half year's time. Only two were retarded.

2. **Alexander**, J. M. "Binet-Simon test in practical use in the public schools of Hinsdale, Illinois," Elementary School Journal, 21: 146-48, October, 1920.

Fifty-three first-grade pupils were tested and rearranged on the basis of their ability in three sections, the best of which did extra work. Also special classes in other grades were tested and handled in accordance with the results.

3. **Almack**, J. C. and J. L. "Administrative problems connected with gifted children," Educational Administration and Supervision, 8:129-36, March, 1922.

In the junior high school of Eugene, Oregon, about 800 pupils were tested with the Army Alpha, Otis and Stanford Revision. Fifty-one with I. Q.'s above 100 were selected for superior sections.

4. **Anderson**, **Rose** G. "Methods and results of mental surveys," Journal of Applied Psychology, 6:1-28, March, 1922.

A general discussion of a number of mental surveys and of the number of pupils selected as feeble-minded. This varied from .16 percent to 6.4 percent.

5. **Armentrout**, W. D. "Classification of junior high-school pupils by the Otis scale," Journal of Educational Psychology, 11:165-68, March, 1920.

Four hundred junior high-school pupils in Lawrence, Kansas, were tested with the Otis scale and three or four groups in each half year selected according to their I.Q.'s.

6. **Armentrout**, W. D. "Grouping pupils by intelligence tests," School Review, 28:249-51, April, 1920.

Another account of procedure in the Lawrence, Kansas, Junior High School.

7. **Armentrout**, W. D. "Classification and promotion of pupils," Education, 42:506-12, April, 1922.

A discussion of several flexible plans. Advocates seven, eight, and nine year courses for the elementary school with classification based upon intelligence, physical and educational tests, school marks and teachers' estimates.

8. **Armentrout**, W. D. "Classification of junior high-school pupils by the Otis scale," Education, 43:83-87, October, 1922.

Another account of the classification of the junior high-school pupils at Lawrence, Kansas.

9. **Arthur**, **Grace**. "An application of intelligence tests to the problem of school retardation," School and Society, 10:614-20, November 22, 1919.

In a St. Paul elementary school, group and individual tests were used and pupils placed according to the results. In two years failures were reduced from 11 to 2.9 percent, special promotions increased from 1.5 to 9.4 percent and $4000 saved.

10. Ayer, F. C. "The present status of promotional plans in city schools," American School Board Journal, 66:37-39, April, 1923.

Gives questionnaire replies from 124 cities. Thirty-six different plans are listed. The cities average about ten plans apiece. It was found that there is little evidence as to the value of the various plans.

11. Badanes, Julie E. "The first practical steps in selecting gifted children in a large city school." New York: Continental Printing Company, 1921. 22p.

A brief history of intelligence testing discussing especially the work and writings of Burt, Meumann, Allen, Pearson and others.

12. Baer, J. A. "Individual differences among pupils," Cleveland, Ohio; Cuyahoga County Public Schools, 1922. 28p.

A general discussion of the individual differences that exist followed by suggestions as to different bases and plans of classification and what the teacher can do to look after individual differences in the different subjects. This is followed by a short bibliography.

13. Bagley, W. C. "Educational determinism; or democracy and the I. Q.," School and Society, 15:373-84, April 8, 1922. Educational Administration and Supervision, 8:257-72, May, 1922.

A strong plea against classifying or instructing pupils according to their I.Q.'s as yielded by mental tests.

14. Bagley, W. C. "Professor Terman's determinism: A rejoinder," Journal of Educational Research, 6:371-85, December, 1922.

An answer to Terman's criticism of Bagley's address on Educational Determinism.

15. Bagley, W. C."Educational determinism again: A rejoinder to Professor Whipple's reply," School and Society, 16:141-44, August 5, 1922.

A further argument against the use of intelligence test results for classifying school children.

16. Barton, J. W. "School organization on an objective basis," Educational Administration and Supervision, 6:187-97, April, 1920.

In Elk River, Minnesota, 323 pupils in grades one to twelve were tested with the Kansas Silent Reading and an opposites test. Thirteen percent were given extra promotion on the basis of test results and of these all but three made good. Not a single one was injured physically.

17. Bates, Grace M. "The work of the students' welfare committee of Erasmus Hall High School," Bulletin of High Points in the Work of the High Schools of New York City, 5:20-24, June, 1923.

In this school, group intelligence test results are used to classify the pupils on the first day of actual high-school work. Whatever adjustments seem advisable are made later but it has been found that the original classification is fairly accurate.

18. Batson, W. H. "The South Dakota group intelligence test for high schools," School and Society, 15:311-15, March 18, 1922.

An account of testing almost 1500 pupils in South Dakota. Shows how test results compare with teachers' estimates, also the amount of variation between pupils and schools.

19. Berry, C. S. "Classification by tests of intelligence of ten thousand first-grade pupils," Journal of Educational Research, 6:185-203, October, 1922.

An account of the plan used in Detroit. First-grade pupils are tested and divided into X, Y and Z groups according to their ratings. The course of study differs for the three groups.

20. Biddle, Anne E. "The superior group," News Letter, 17:3-4, May, 1923.

A brief account of grouping by ability in the high schools of a certain city during the last several years with individual accounts of a number of bright pupils. Superior pupils were selected by the judgment of the teachers verified by intelligence tests. The writer is heartily in favor of the plan.

21. Bliss, D. C. "The application of standard measurements to school administration," Fifteenth Yearbook of the National Society for the Study of Education, Part I. Bloomington, Illinois: Public School Publishing Company, 1916, p. 69-77.

A general discussion of the question with some illustrations drawn from Montclair, New Jersey.

22. Book, W. F. "Variations in mental ability and its distribution among the school population of an Indiana county," Fifth Conference on Educational Measurements, Bulletin of the Extension Division, Vol. 4, No. 4. Bloomington: Indiana University, 1918, p. 100-31.

An account of the use of Indiana Schedule B in all the schools of a county. Discusses the differences found between pupils, the distribution of ability, the reliability of the scale used and the adjustments that should be made in the schools.

23. Bracewell, R. H. "The Freeman-Rugg general intelligence tests as an aid to economy in school administration," School Review, 29:460-66, June, 1921.

In the Burlington, Iowa, High School, pupils were tested with these tests and the average of the test results and teachers' rankings used for purposes of placement.

24. Branson, E. P. "An experiment in arranging high-school sections on the basis of general ability," Journal of Educational Research, 3:53-55, January, 1921.

In the Long Beach, California, High School, freshmen were grouped in English according to their scores on the Otis scale. This scale appeared to select inferior pupils better than superior ones, altho it was fairly efficient in both respects.

25. Breed, F. S. "Shall we classify pupils by intelligence tests?" School and Society, 15:406-9, April 15, 1922.

A general discussion listing a number of principles, cautions, conditions, and steps in using intelligence tests. The conclusion is that altho intelligence test results, the will, and the emotions, etc., are important factors in classifying, the achievement secured must be the final determining factor.

26. Breed, F. S. and Breslich, E. R. "Intelligence tests and the classification of pupils. II." School Review, 30:210-26, March, 1922.

An account of an experiment in the University of Chicago High School in which it was found that Otis test results divided pupils better than measures of their ability in arithmetic. The conclusion reached was that no test gives a reliable basis for permanent classification.

27. Brooks, S. S. "Some uses for intelligence tests," Journal of Educational Research, 5:217-38, March, 1922. Improving Schools by Standardized Tests. Boston: Houghton Mifflin Company, 1922. p. 96-123.

Recommends that pupils be placed by intelligence test results rather than by those from achievement tests.

28. Buckingham, B. R. "Suggestions for procedure following a testing program-I. Reclassification," Journal of Educational Research, 2:787-801, December, 1920.

States that intelligence tests alone are not a safe guide for promotion but should be supplemented by subject-matter tests and measures of non-intellectual qualities.

29. Burkard, W. E. "Grouping by ability in the Robert Treat Junior High, Newark," News Letter, 19:4-5, April, 1921.

In Newark there were accelerated, normal and retarded groups in grades I-IX based on measurements by a modified form of the Binet Scale. The same course of study was followed but at different rates. The plan appears to have given satisfaction.

30. Callihan, T. W. "An experiment in the use of intelligence tests as a basis for proper grouping and promotions in the eighth grade," Elementary School Journal, 21:465-69, February, 1921.

Almost 300 eighth-grade pupils were divided into sections according to ability. The basis was a combination of results from the Illinois General Intelligence Scale, Monroe's Silent Reading Test, and to some extent teachers' estimates and marks.

31. Campbell, Cora. "Intelligence tests as a basis for classification," The Technique of Supervision by the Elementary School Principal. The First Yearbook of the Department of Elementary School Principals. Washington: National Education Association, May, 1922, p. 45-49.

As a result of a number of experiments Kansas City adopted the policy of classifying pupils upon the basis of individual intelligence tests given in the kindergarten and first grade, group tests given in other grades, and health, attendance, school attitude and certain other factors.

32. Campbell, L. H. "Age gradation and grade grouping," American School Board Journal, 58:36, May, 1919.

States that in Providence, Rhode Island, pupils were grouped according to their age and placings verified by the Stanford Revision. Ten good results are listed.

33. Carback, Clarence. "Grouping of children by abilities and consequent change in school procedure. 3. Procedure in Philadelphia." Ninth Annual Schoolmen's Week Proceedings. University of Pennsylvania Bulletin, Vol. 23, No. 1, p. 269-72.

States that in Philadelphia there are about twenty elementary schools with some or all pupils grouped according to their abilities. The bright pupils take an enriched curriculum.

34. Cayco, F. and Pressey, S. L. "Three refinements of method in school surveys," Educational Administration and Supervision, 7:433-38, November, 1921.

A discussion of the need of relating age-grade, achievement and intelligence test data to each other. Also a brief account of testing the pupils of a small system.

35. Chassell, C. S. and Chassell, L. M. "A survey of the three first grades of the Horace Mann school by means of psychological tests and teachers' estimates and a statistical evaluation of the measures employed," Journal of Educational Psychology, 12:72-81, 243-52; February, May, 1921.

The average of results according to the Stanford Revision, Pressey and Myers Tests and teachers' estimates of maturity and reading ability was found and used as a basis of forming three fairly homogeneous groups.

36. Clerk, F. E. "Providing for individual differences by grouping by abilities. Organization and practical working of the plan." Eighth Annual Schoolmen's Week Proceedings, University of Pennsylvania Bulletin, Vol. XXI, No. 37, Philadelphia, June 18, 1921, p. 243-49.

An account of procedure in Winchester, Virginia, in which teachers' judgments, school marks, achievement and intelligence tests, etc., are used in placing pupils. Accelerated, normal and slow groups are formed in high school as well as grades. A number of cautions are given.

37. Cleveland, Elizabeth. "Detroit's experiment with gifted children," School and Society, 12:179-83, September 11, 1920.

The best five percent of seventh and eighth-grade pupils are selected and placed in special classes. The first basis of selection was the teacher's judgment, later the Binet Tests were used and finally group tests. The pupils selected gain a little time and take an enriched course.

38. Cleveland, Elizabeth. "Some further studies of gifted children," Journal of Educational Research, 4:195-99, October, 1921.

A study of 144 bright and the same number of average children in Detroit. Health, home conditions, tastes, etc., were studied.

39. Cole, L. W. "Mental age and school entrance," School and Society, 8:418-19, October 5, 1918.

In Denver pupils entering the first grade are tested. It is found that the I.Q. of five-year-olds is high, of six-year-olds next and that of seven-year-olds lowest. The conclusion is that better results could be obtained by selecting new entrants by the Binet Scale than by chronological age.

40. Cole, L. W. "Prevention of the lockstep in schools," School and Society, 15:211-17, February 25, 1922.

An account of the use of the Binet, and Cole and Vincent Tests in the first grade. It is shown that six-year-old children are not a homogeneous group and that there is a great difference between the best and worst entering the first grade. States that classes should be organized on a mental-age basis.

41. Colvin, S. S. "The present status of mental testing," Educational Review, 64:196-206, 320-37; October, November, 1922.

A brief account of the testing movement and present knowledge concerning it and a discussion of the use of tests for the classification of pupils, their validity, cautions to be observed in their use, etc.

42. Colvin, S. S. et al. "Intelligence tests and their use," Twenty-first Yearbook of the National Society for the Study of Education. Bloomington, Illinois: Public School Publishing Company, 1922, 289 p.

The most complete discussion of the subject published. Part I deals with the nature, history and general principles of intelligence testing and Part II with the administrative use of intelligence tests. The latter part contains both theoretical discussion and actual accounts of what is being done.

43. Cox, E. M. "Report of committee on promotions and rates of progress." Superintendent's Report, Oakland, California, 1917-18, p. 77-79.

History of acceleration, retardation, special progress, mental testing, special classes, etc., in Oakland. States that the results of special promotions, classes and mental testing have been good.

44. Cox, P. W. L. "Providing for individual differences by means of grouping by ability." Ninth Annual Schoolmen's Week Proceedings, University of Pennsylvania Bulletin, Vol. 23, No. 1, p. 233-44.

An account of grouping pupils in junior and senior high schools according to results from the Otis Test, various achievement tests and their interests and aptitudes.

45. Coxe, W. W. "School variation in general intelligence," Journal of Educational Research, 4:187-94, October, 1921.

Gives results of giving the Otis Test in twenty-four Cincinnati elementary schools as a basis for selecting pupils for a special six-year classical course. Great variation in intelligence was found.

46. Cummins, R. A. "Educational measuring sticks and their uses," American School Board Journal, 63:33-5, August, 1921.

Contains a discussion of the use of intelligence and subject-matter tests.

47. Dawson, C. D. "Classification of kindergarten children for first grade by means of the Binet scale," Journal of Educational Research, 6:412-22, December, 1922.

In Grand Rapids 64 percent of 2000 kindergarten children were classified in the first grade on the basis of test results. Three groups were formed. Eighteen percent had to be shifted from the original classification.

48. Deamer, Arthur. "An experiment in acceleration," Fargo, North Dakota; Board of Education, 1919, 32 p.

Pupils for fast sections in the upper grades were selected according to their school marks. When measured with standardized tests they showed that their achievement was satisfactory. Likewise their high-school records were good.

49. Dickson, V. E. "The relation of mental testing to school administration." The Normal Seminar Bulletin A, No. 1. Cheney, Washington, June, 1917.

A study of the differences of pupils according to sex, chronological and mental age, I.Q., school mark, and teacher's estimate. The study shows a high correlation between mental ages and school marks.

50. Dickson, V. E. "What first-grade children can do in school as related to what is shown by mental tests," Journal of Educational Research, 2:475-80, June, 1920.

In Oakland, California, first-grade pupils are classified in three groups on the basis of test results and school marks. Pupils with a mental age below six cannot do regular first-grade work and should be put in a special group.

51. Dickson, V. E. "The use of group mental tests in the guidance of eighth-grade and high-school pupils," Journal of Educational Research, 2:601-10, October, 1920.

On the basis of scores on part of the Otis Scale and school marks an accelerated class was formed which covered eighth-grade work in fifty-five days, entered high school immediately and made better marks than the average. Two similar groups are also discussed.

52. Dickson, V. E. "The use of mental tests in school administration," Monograph, Vol. IV. Berkeley, California: Board of Education, June, 1922. 44 p.

Advocates the use of test results in determining the grade placement of pupils. Shows that individual intelligence tests are rather highly reliable and group tests fairly so, also that the I.Q. is a fairly safe measure of prognosis.

53. Dickson, V. E. "The treatment of gifted children in Oakland and in Berkeley." Proceedings of the First Annual Conference on Educational Research and Guidance, San Jose State Teachers College. Sacramento, California: State Printing Office, 1923, p. 26-28.

A discussion of the basis of segregation and what should be done after segregation. Argues that children should be treated differently according to their differences in ability. Also that gifted children should not merely gain time but should also have a broader and surer foundation.

54. Dickson, V. E. Mental Tests and the Classroom Teacher. Yonkers-on-Hudson: World Book Company, 1923. 231 p.

This is, as its name indicates, a general discussion of the use of mental tests in school. The author maintains that group tests are fairly reliable for predicting school success and that pupils should be classified at the time of entering school on the basis of their ability. The results of mental tests should be an important factor in this classification.

55. Doten, Willard. "Tests for ability grouping," American School Board Journal, 63:37-38, October, 1921.

The Montclair, New Jersey Junior High School pupils were grouped according to the results from the National and Haggerty Tests. The work for the groups was differentiated. Satisfactory results followed.

56. Edmondson, Margaret B. "A mental survey of first-grade school pupils," Pedagogical Seminary, 27:354-70, December, 1920.

A discussion of the mental ages, I.Q.'s, school marks, chronological ages, and teachers' estimates of some grade pupils in Eugene, Oregon. Shows the lack of homogeneity of the group and the relationship between the various data.

57. Ettinger, W. L. "Economy in school administration," School and Society, 14:409-15, November 12, 1921.

An argument that much of the waste in school is due to crude classification.

58. Ettinger, W. L. "Crude classification causes waste," School Life, 7:66, November, 1921.

A general statement of the need of careful classification.

59. Fordyce, Charles. "Intelligence tests and classifying children in the elementary school," Journal of Educational Research, 4:40-42, June, 1921.

An account of the use of the Haggerty Test on over 1000 pupils. Discusses relation of test results and teachers' estimates.

60. Frank, C. D. "A ray of light let in by a study of first-term failures," Bulletin of High Points in the Work of the High Schools of New York City, 5:6-7, May, 1923.

An account of an experiment showing that elementary-school marks furnish a better basis of predicting high-school marks than do intelligence test results.

61. Freeman, F. N. "Bearing of the results of mental tests on the development of the child," Scientific Monthly, 12:558-70, June, 1921.

States that the traditional method brings together pupils of very different mental ability, while grouping by mental ability brings together pupils whose intellectual processes are not equally developed. Ideal grouping would bring together pupils alike in both.

62. Freeman, F. N. "Bases on which students can be classified effectively," School Review, 29:735-45, December, 1921.

States that classification should be based upon mental ability, chronological and physical age, social and intellectual stages of development. Both tests and teachers' judgments should be used.

63. Frasier, G. W. "The measurement of intelligence as an aid to administration," Educational Administration and Supervision, 6:361-66, October, 1920.

An account of administering the Stanford Revision in Spokane and giving pupils special promotions and demotions according to the results. In general the placement of pupils was satisfactory.

64. Glass, J. M. "Classification of pupils in ability groups," School Review, 28:495-508, September, 1920.

In a Rochester junior-high school pupils were placed by results from several intelligence tests. Most of those so placed did satisfactory work but a few had to be shifted.

65. Goddard, H. H. "Two thousand children tested by the Binet measuring scale for intelligence," Addresses and Proceedings of the National Education Association, 49:870-78, 1911.

An account of using the Binet scale to find New York children three years or more below normal.

66. Goldstone, G. A. "Differentiation of method in teaching reading to slow and bright pupils," Bulletin of High Points in the Work of the High Schools of New York City, 5:11-14, April, 1923.

An account of an experiment in grouping high-school pupils into bright, medium and slow groups according to the results of the Thorndike-McCall Reading Scale. After the groups were formed each was taught by a different method but with practically the same subject matter. The results were satisfactory.

67. Gray, P. L. and Marsden, R. E. "An application of intelligence tests," Journal of Experimental Pedagogy and Training College Record, 6:33-38, March 5, 1921.

An account of testing a small group of pupils with the Stanford Revision. Compares the results with those from examinations, teachers' estimates and vocabulary tests.

68. Greenberg, B. B. "Intelligence tests as a basis for reclassification," The Technique of Supervision by the Elementary School Principal. The First Yearbook of the Department of Elementary School Principals. Washington: National Education Association, May, 1922, p. 55-58.

An account of the reorganization of a New York school on the basis of results from individual and group intelligence tests, several achievement tests, sensory and motor tests, etc. Reclassification on the basis of these tests resulted in increased homogeneity in each grade.

69. Haggerty, M. E. "Recent developments in measuring human capacities," Journal of Educational Research, 3:241-53, April, 1921.

States that intelligence must be supplemented by other traits as a basis for estimating capacities.

70. Hanson, W. A. "Mental measurement and the special class system in New London, Connecticut," American School Board Journal, 63:37, 119, September, 1921.

Twenty-seven hundred children were tested with the Dearborn and Haggerty Tests. Special classes were formed and the results also used to guide teachers in their instruction. In some cases individual tests were given.

71. Harper, May M. "Intelligence tests as a basis for homogeneous grouping," Elementary School Journal, 22:781-82, June, 1922.

In a junior high school in Xenia, Ohio, pupils were classified in three groups according to the Haggerty and Terman tests. Reclassifications were made at the end of every two months. About half of the pupils were changed at the end of two months and about one-fourth at the end of four months.

72. Henmon, V. A. C. "The measurement of intelligence," School and Society, 13:151-58, February 5, 1921.

An argument in favor of the cautious use of tests for purposes of classification. States that these tests should measure social and mechanical as well as abstract intelligence.

73. Henry, Mary B. "Mental testing as an aid in guidance and classification of school children." Department of Research of the Santa Ana Public Schools, Bulletin No. 1. Santa Ana, California: Board of Education, 1919. 23 p.

A general account of the use of mental tests in the Santa Ana schools. Discusses their agreement with teachers' estimates, their use for the purpose mentioned above, etc. Pupils are grouped in three sections on the basis of test results and other information.

74. Herrig, Anna B. "Promotions in the practise school as determined by the use of standard tests and educational measurements," Educational Administration and Supervision, 7:217-25, April, 1921.

A theoretical discussion.

75. Hines, H. C. "What Los Angeles is doing with the results of testing," Journal of Educational Research, 5:45-47, January, 1922.

Gives various details connected with the use of both intelligence and achievement tests in Los Angeles.

76. Hines, H. C. "Measuring the intelligence of school pupils," American School Board Journal, 64:35-37, 135, April, 1922.

A rather good theoretical discussion giving criticisms on both sides of the question and listing what are considered the best tests for the different grades.

77. Holley, C. E. "Mental tests for school use." Bureau of Educational Research Bulletin, No. 4, University of Illinois Bulletin, Vol. 17, No. 28, March 8, 1920. University of Illinois, 91 p.

A study of the use of a number of mental tests with some theoretical discussion.

78. Holmes, H. W. "The general philosophy of grading and promotion with relation to intelligence testing," School and Society, 15:457-61, April 29, 1922.

An argument for an enriched course for superior pupils.

79. Hughes, W. H. "Provisions for individual differences in high-school organization and administration," Journal of Educational Research, 5:62-71, January, 1922.

A questionnaire study of plans actually used by high schools.

80. Hunter, F. N. "Report of the superintendent of schools for the year 1917-18." Oakland, California: Board of Education, 1919. 353 p.

Contains considerable discussion of the Otis and Simon-Binet Tests, flexible promotion, etc.

81. Irving, M. Louise. "Classification into ability groups in Santa Rosa," Journal of Educational Research, 6:362-64, November, 1922.

In the junior high school of Santa Rosa classification is based upon two or three intelligence tests and several achievement tests.

82. Jahrling, Robert. "Educating gifted children in Hamburg," Pedagogical Seminary, 30:35-39, March, 1923.

Gifted children are selected by a combination of tests, teachers' estimates and school marks. A special course is provided for them.

83. **Johnson, O. J.** "Teachers' estimates of qualities of gifted pupils as related to classroom activities," School and Society, 17:466-69, April 28, 1923.

Thirty-three teachers of 900 pupils, mostly in high school, answered a questionnaire of ten questions on the general topic given above. The teachers favored separating bright pupils from others and giving them special work.

84. **Jordan, R. H.** "An example of classification by group tests," Educational Administration and Supervision, 6:198-201, April, 1920.

In a Minnesota grade school the seventh and eighth-year pupils were tested with several mental tests. Those above the 75 percentile of the next higher class were considered for extra promotion, those below the 25 percentile of the next lower class for failure or demotion.

85. **Keener, E. E.** "The use of measurements in a small city school system," Journal of Educational Research, 3:201-06, March, 1921.

This article mentions a few instances of readjustment in school based upon test results. The averages given are from the Richmond, Indiana, schools.

86. **Kelley, T. L.** "Again: educational determinism," Journal of Educational Research, 8:10-19, June, 1923.

The writer states that he is in sympathy with Bagley's belief in democracy and Terman's insistence upon the reality of individual differences. His conclusion is that instruction should be properly differentiated.

87. **Kent, R. A.** "An experiment in the grading and placing of children," American School Board Journal, 62:40-41, February, 1921.

In Duluth, pupils were placed according to results from the Otis test. Discipline was made easier, interest was increased, better work was done and other good results secured.

88. **Ketner, Sarah P.** "Grouping by standardized tests for instructional purposes," Journal of Educational Research, 2:620-25, October, 1920.

A Denver school grouped pupils in reading, arithmetic and penmanship according to the results of tests. The experiment seemed to be successful and the plan was to be extended to other schools.

89. **Kuhlmann, F.** "What constitutes feeble-mindedness," Journal of Psycho-asthenics, 19:214-36, June, 1915.

Reaches the conclusion that the I.Q. is the most reliable criterion of an individual's grade of intelligence.

90. KUNTZ, E. E. "Grouping of children by abilities and consequent changes in school procedure. 3. Procedure in Lansford." Ninth Annual Schoolmen's Week Proceedings, University of Pennsylvania Bulletin, Vol. 23, No. 1, p. 267-69.

The Haggerty, National and Terman Tests were used as a basis for grouping pupils. Two or three sections were formed in each of the lower grades and more time given to slow pupils in the upper grades.

91. KYTE, G. C. "An experiment in the education of gifted children in the first grade," The Technique of Supervision by the Elementary School Principal. The First Yearbook of the Department of Elementary School Principals. Washington: National Education Association, May, 1922, p. 71-80.

In a public school at Berkeley superior first-grade children are selected on the basis of results from the Stanford Revision and certain other data and allowed to gain time. A reading test showed that nothing was lost in achievement. On the whole the plan was satisfactory.

92. LAWS, A. R. and BOWIE, S. "Intelligence tests in examinations for junior scholarships," Journal of Experimental Pedagogy and Training College Record, 6:155-69, December 5, 1921.

Found that the correlation between intelligence test results and terminal marks ranged from —.04 to +.96. Concludes that no one test is sufficient as a basis for judging pupils.

93. LAYTON, W. K. "The intelligence testing program of the Detroit public schools," School and Society, 15:368-72, April 1, 1922.

Much the same as contained in the reference by same author in the Twenty-first Yearbook. Especially describes the X, Y and Z groups in the first grade, membership in which is determined by test results and which have different work.

94. LINDSAY, M. D. and GAMSBY, R. S. "Where test scores and teachers' marks disagree," School Review, 29:679-87, November, 1921.

An account of the use of the Terman Test in Palo Alto Union High School. It was found that in all cases of poor work by superior students the teachers could have secured better work if they had known the pupils' ability.

95. LLOYD, S. M. and ULLRICH, O. A., JR. "The progress of pupils in an ungraded class," Psychological Clinic, 11:276-87, February 15, 1918.

Several subject-matter tests were used with the pupils of an ungraded room in Austin. Pupils gained by being placed in this room.

96. LOWELL, FRANCES. "An experiment in classifying primary grade children by mental age," Journal of Applied Psychology, 6:276-90, September, 1922.

In the practise school of the City Normal School of Rochester, pupils of grades I to III were grouped with the Kuhlmann Revision as a chief basis.

97. Madsen, I. N. "Interpreting achievement in school in terms of intelligence," School and Society, 14:59-60, July 30, 1921.

A rather short discussion of the E. Q., A. Q. and I. Q.

98. Madsen, I. N. "Intelligence as a factor in school progress," School and Society, 15:283-88, March 11, 1922. American School Board Journal, 64:37-38, June, 1922.

A number of towns, mostly in Idaho, gave the Haggerty Test in the elementary school and the Army Alpha in high school to over 12,000 pupils. It was found that the I.Q.'s of the same age group varied directly with the grade placement.

99. Marshall, Jessica. "Using the results of testing," The Technique of Supervision by the Elementary School Principal. The First Yearbook of the Department of Elementary School Principals. Washington: National Education Association, May, 1922, p. 49-55.

An account of the use of tests for purposes of placing pupils in a Toledo school. In the lower grades classification was chiefly according to the results of various intelligence tests, in the upper grades according to those from achievement tests.

100. Martens, Elise H. "Better classification," The Journal of the National Education Association, 12:174-75, May, 1923.

A statement of the problem and our failure to meet it at present followed by an argument that individual differences must be recognized. It is implied that mental and subject-matter tests should be used for this purpose.

101. Miller, W. S. "General intelligence tests," School Review, 28:94, February, 1920.

A brief statement that the University of Minnesota High School has been classifying entering students according to mental test results with a list of favorable results. States that the highest correlation between mental test results and achievement is in mathematics and science.

102. Mitchell, David. "Psychological examination of pre-school age children," School and Society, 15:561-68, May 20, 1922.

Discusses the results of testing 1000 New York children, advocates special classes, differentiated curricula, etc.

103. Montgomery, E. W. "Group tests for intelligence in the Bedford (Indiana) schools," Sixth Conference on Educational Measurements, Bulletin of the Extension Division, Vol. 5, No. 1. Bloomington: Indiana University, 1919, p. 54-55.

The Binet-Simon Test was used to select pupils for special classes. In addition group tests were used in high school as the basis of organizing fast sections.

104. Murdoch, Katharine. "Rate of improvement of the feeble-minded as shown by standardized educational tests," Journal of Applied Psychology, 2:243-49, September, 1918.

The Stanford Revision and a dozen or so subject matter tests were used on a small number of feeble-minded children. Conclusion reached was that the mental age of a feeble-minded is not equivalent to the same mental age of a normal child.

105. Myers, A. F. "Reclassification of children on basis of tests in Port Clinton schools," Journal of Educational Method, 1:24-25, September, 1921.

In Port Clinton, Ohio, a change was made from the annual to the semester system by promoting the upper 40 percent of each grade one semester and demoting the lower 10 percent one semester. The basis of classification was a combined intelligence and achievement test score.

106. Myers, C. E., Myers, G. C. and Layton, S. H. "Group mental testing in Altoona, Pennsylvania," School and Society, 13:624-28, March 28, 1921.

The Myers' Mental Measure was used. It was recommended that special classes for sub- and super-normal children be formed, that several parallel courses be organized, that pupils be classified according to their I.Q.'s and school marks, etc.

107. Myers, G. C. "Economy in intelligence classification," Educational Administrtion and Supervision, 6:309-12, September, 1920.

A theoretical discussion of the grouping of pupils, maximum and minimum curricula, etc. Also an account of the testing and grouping of 1500 illiterate soldiers into four groups in each grade.

108. Myers, G. C. "Intelligence classification and mental hygiene," Pedagogical Seminary, 28:156-60, June, 1921.

An argument for complete mental surveys of school systems.

109. Nichols, M. L. "The teaching of science to classes divided according to ability." Ninth Annual Schoolmen's Week Proceedings, University of Pennsylvania, Vol. 23, No. 1, p. 324-26.

In the South Philadelphia High School for Girls science classes were divided according to results of intelligence tests. Different methods of teaching were used in the different sections.

110. Omans, A. C. "Provision for ability grouping in junior and senior high school," American School Board Journal, 65:55-58, 138, October, 1922.

Gives questionnaire replies from fifty schools in all parts of the country. Shows that in general grouping on an ability basis is favored over individual promotions. Slightly over one-half classified on intelligence tests and teachers' estimates, about two-fifths on intelligence tests and school marks and a few on school marks alone. This is a good discussion and summary of practises.

111. PATTERSON, D. G. "A mental survey of the school population of a Kansas town," School and Society, 7:84-89, January 19, 1918.

This is well described by the title. A comparison is also made with Pintner's results in Ohio.

112. PINTNER, R. "A mental survey of the school population of a village," School and Society, 5:597-600, May 19, 1917.

An account of the survey of an Ohio village with a group of tests.

113. PINTNER, R. The Mental Survey. New York: D. Appleton and Company, 1918, 116 p.

This contains an account of using a number of intelligence and achievement tests, the results obtained from surveying school systems and the relation between educational accomplishments and mental ability, etc.

114. PINTNER, R. and CUNNINGHAM, BESS V. "The problem of group intelligence tests for very young children," Journal of Educational Psychology, 13:465-72, November, 1922.

A brief account of the organization of three sections in grade I of a school at Toledo, Ohio. The Pintner-Cunningham Primary Tests were used.

115. PINTNER, R. and MARSHALL, HELEN. "A combined mental-educational survey," Journal of Educational Psychology, 12:32-43, January, 1921.

A discussion of the necessity of combining the results of mental and educational tests. Also states that the schools have been best adapted to dull pupils.

116. PINTNER, R. and MARSHALL, HELEN. "Results of the combined mental-educational survey tests," Journal of Educational Psychology, 12:82-91, February, 1921.

A comparison of school work with capacity by means of mental and educational indices.

117. PINTNER, R. and NOBLE, H. "The classification of school children according to mental age," Journal of Educational Research, 2:713-28, November, 1920.

About 5000 pupils in a Columbus, Ohio, school were tested with the Stanford Revision. Fifteen percent were either given extra promotion or demoted. Special classes and an ungraded class were formed, acceleration and retardation were both reduced, discipline, work and interest were bettered.

118. PRESSEY, L. C. "The relation of intelligence to achievement in the second grade," Seventh Conference on Educational Measurements, Bulletin of the Extension Division, Vol. 6, No. 1. Bloomington: Indiana University, 1920, p. 68-77.

Describe the use of the Indiana Primer Scale and Scale of Attainment No. I. in an Indiana city. It was found that no inferior pupils were high in achievement but that many superior pupils were low.

119. Pressey, S. L. "A systematic plan for selecting sub-normal and super-normal children in the public schools," Fifth Conference on Educational Measurements, Bulletin of the Extension Division, Vol. 4, No. 4. Bloomington: Indiana University, 1918, p. 92-99.

Recommends that a group test be used first and then the Stanford Revision.

120. Pressey, S. L. "The 'efficiency' of a group scale of intelligence in prognosticating success and failure in the junior high school," Journal of Applied Psychology, 3:381-85, December, 1919.

Discusses the use of the Indiana Group Point Scale in Bloomington. States that results are rather highly correlated with failure and success in school work.

121. Pressey, S. L. "School surveys by means of group tests of intelligence," Sixth Conference on Educational Measurements, Bulletin of the Extension Division, Vol. 5, No. 1. Bloomington: Indiana University, 1919, p. 46-53.

Discussion of the use of the Indiana Cross-Out Scale which is useful for general survey purposes.

122. Pressey, S. L. "Suggestions with regard to the use of mental tests and in particular with regard to their use in combination with tests of achievement," Seventh Conference on Educational Measurements, Bulletin of the Extension Division, Vol. 6, No. 1. Bloomington: Indiana University, 1920, p. 78-80.

A discussion of general mental surveys, the use of achievement tests, their prognostic value, etc.

123. Pressey, S. L. "An attempt to measure the comparative importance of general intelligence and certain character traits in contributing to success in school," Elementary School Journal, 21:220-29, November, 1920.

A study of junior high-school pupils. The correlation of health, attitude, preparation and ability with school marks and with age was studied. Ability and attitude were found to be the most important.

124. Pressey, S. L. and L. W. "Measuring the 'usefulness' of tests in solving school problems," School and Society, 12:531-34, November 27, 1920.

The Indiana Cross-Out Scale was given to pupils in the upper grades and fast and slow sections arranged according to its results. Likewise the Primer Scale was given in the first grade. A better agreement was found between placement by test results and semester marks than between placement by teachers' judgments and semester marks.

125. Price, E. D. "The Enid plan of classification of pupils according to mental ability." Enid, Oklahoma: Board of Education, 1921. 12 p.

Elementary pupils were grouped into three groups according to teachers' opinions. The superior and inferior groups were also tested by the Terman Test, the results of which usually verified the teachers' judgment.

126. Proctor, W. M. "The use of intelligence tests in the educational guidance of high-school pupils," School and Society, 8:473-78, 502-09, October 19, 26, 1918.

In Palo Alto High School it was found that the Stanford Revision apparently affords as good a measure of success in school as do teachers' estimates or previous school marks. Pupils with I. Q.'s of 95 or below are not likely to be successful in high-school work.

127. Proctor, W. M. "Psychological tests and the probable school success of high-school pupils," Journal of Educational Research, 1:258-70, April, 1920.

A discussion of the use of the Stanford Revision and the Army Alpha Tests in high school and the connection between intelligence and high-school marks, elimination, retardation, etc. Army Alpha yields a rather low correlation with school marks.

128. Proctor, W. M. "The use of psychological tests in the educational guidance of high-school pupils," Journal of Educational Research, 1:369-81, May, 1920.

Shows that pupils guided in the choice of subjects according to Stanford Revision results show less elimination and failure than do those who are not so guided.

129. Proctor, W. M. "The use of psychological tests in the vocational guidance of high-school pupils," Journal of Educational Research, 2:533-46, September, 1920.

A discussion of the occupational levels of intelligence as found by testing in the army, the vocational choices of high-school pupils and the connection between the two. States that pupils should be guided out of vocations that require a greater degree of intelligence than they possess.

130. Proctor, W. M. and Ward, Helen. "Relation of general intelligence to the persistence of educational and vocational plans of high-school pupils," Journal of Educational Research, 7:277-88, April, 1923.

Shows that there is a definite relation between results of intelligence tests and vocational plans and persistence and a still higher relation between the former and educational plans and persistence. Implies that groups of pupils formed according to intelligence test results should receive different treatment in our schools.

131. Race, Henrietta V. "A study of a class of children of superior intelligence," Journal of Educational Psychology, 9:91-98, February, 1918.

In a Louisville school, pupils with I. Q.'s above 120 were selected. They did one year's work in one-half year without over twenty minutes of home study daily. The pupils appeared to be greatly benefited by this work.

132. Rapp, Anna A. "Grouping of children by abilities and consequent changes in school procedure. 2. Procedure in Reading." Ninth Annual Schoolmen's Week Proceedings, University of Pennsylvania Bulletin, Vol. 23, No. 1, p. 263-67.

The Terman and Illinois Tests were used in the seventh and eighth grades and pupils classified according to scores.

133. Reeve, W. D. "Homogeneous grouping of high-school students by means of psychological tests," Fourth Yearbook of the National Association of Secondary School Principals. Menasha, Wisconsin: George Banta Publishing Company, 1921, p. 81-94.

States that it is a waste of time to attempt to teach the best and worst 10 percent along with average pupils. Discusses the correlation of tests and school marks and says that classification by test results makes it easier to provide suitable instruction, reduces failures, gives the brighter pupils better training and in general conserves human resources.

134. Robbins, Chester. "The initial grouping of pupils." Ninth Annual Schoolmen's Week Proceedings, University of Pennsylvania Bulletin, Vol. 23, No. 1, p. 244-48.

A discussion of various articles in books dealing with this subject.

135. Rogers, Agnes L. "Mental tests as a means of selecting and classifying college students," Journal of Educational Psychology, 11:181-92, April, 1920.

Discusses results of testing about 300 Goucher college students with the Thorndike and Rogers Tests. Correlations with school marks were low for the single tests but high when the test results were combined. Shows that a division of students into three groups by test results would misplace about one-half, whereas by a chance placement two-thirds would be misplaced.

136. Rowland, S. V. "Individual difference among elementary grade pupils as evidenced by group intelligence tests." Eighth Annual Schoolmen's Week Proceedings, University of Pennsylvania Bulletin, Vol. 21, No. 37, Philadelphia, June 18, 1921, p. 101-04.

An account of classification based on results of the Stanford Revision and the Dearborn Tests. Special classes for inferior children were formed, also two or three groups of those remaining, and some individual instruction given.

137. Ruch, G. M. "An experiment with forced promotion," Educational Administration and Supervision, 6:71-73, February, 1920.

On the basis of several achievement tests a few pupils in the University of Oregon High School were skipped ahead a semester. Most of them made good.

138. Ruch, G. M. "Study of the mental, pedagogical and physical development of the junior division of the University High School, Eugene, Oregon." University of Oregon Publication, Vol. 1, No. 7, 1920.

A rather good discussion along the lines indicated.

139. Ruml, Beardsley. "Reliability of mental tests in the division of an academic group," Psychological Monographs, Vol. 24, No. 4, October, 1917.

A rather lengthy criticism of the use of marks as measures of ability and of the value of a number of tests for dividing pupils into instructional groups.

140. Saam, Theodore. "Intelligence testing as an aid to supervision," Elementary School Journal, 20:26-32, September, 1919. Addresses and Proceedings of the National Education Association, 57:625-29, 1919.

Twenty-three hundred and sixty lower-grade pupils at Council Bluffs, Iowa, were tested with the Stanford Revision. Kindergarten pupils were promoted to first grade on the basis of results. This procedure was satisfactory as a high correlation was found between I.Q.'s and school marks.

141. Schutte, T. H. "A mental survey in the training department of the Moorehead State Teachers College, Moorehead, Minnesota," American School Board Journal, 65:45-47, 53-54, 42-43; October, November, December, 1922.

The Otis, Chicago and Army Alpha Tests were given to elementary pupils and rapid progress groups formed in each grade. Most of the pupils in these groups gained one semester in the year.

142. Seashore, C. E. "Sectioning classes on the basis of ability," School and Society, 15:353-58, April 1, 1922.

Presents the following plan which was found successful in college. Two or three competitive exercises are set up at the first of the year and the marks thereon supplemented by mental tests and high-school marks as a basis of sectioning. Discusses advantages and objections.

143. Seashore, C. E. et al. "Mentality tests; a symposium," Journal of Applied Psychology, 7:229-40, 278-86, 348-60; April, May, June, 1916.

A general theoretical discussion of mental tests, their significance and use.

144. Shideler, J. W. "A correlation of teachers' grades and the scores of intelligence tests," School Review, 29:733-34, December, 1921.

An account of giving the Terman Group Test to 170 high-school pupils at Fort Scott, Kansas. Discusses its correlation with marks in the various subjects and finally concludes that intelligence test results should be used to supplement the teacher's judgment.

145. Spain, C. L. "Grouping of children by abilities and consequent changes in school procedure. 1. Procedure in Detroit." Ninth Annual Schoolmen's Week Proceedings, Vol. 23, No. 1, p. 257-63.

An account of the Detroit plan.

146. Specht, Louise F. "A Terman class in Public School No. 64, Manhattan," School and Society, 9:393-98, March 29, 1919.

Pupils in grades IVB to VIB were selected on the basis of low chronological age, high school marks, I.Q.'s of 120 or better, etc., and allowed to take an enriched curriculum and make individual progress. From one to four grades were covered in six months.

147. Stebbins, R. and Pechstein, L. A. "Quotients, I, E and A," Journal of Educational Psychology, 13:385-98, October, 1922.

A discussion of how to use test results and what should be the basis of classifying pupils.

148. Stetson, P. C. "Homogeneous grouping in the first year of a five-year high school," School Review, 29:351-65, May, 1921.

In the Muskegon, Michigan High School the Chicago Test was found unsatisfactory as a basis of classifying pupils but when supplemented by teachers' opinions gave better results. Both teachers and pupils favored having three groups on the basis of ability.

149. Taylor, J. S. "Grading and promotion," School and Society, 17:405-09, April 14, 1923.

Tells of classification in one New York City district which was based upon I.Q.'s, A.Q.'s, teachers' estimates, etc.

150. Terman, L. M. "Intelligence tests as a basis for grading." The Measurement of Intelligence. Boston: Houghton Mifflin Company, 1916, p. 16-17.

A brief statement that promotions should be made chiefly on the basis of intellectual ability.

151. Terman, L. M. "The use of intelligence tests in the army," Psychological Bulletin, No. 15, p. 177-87, June, 1918.

A good general account of the army work.

152. Terman, L. M. The Intelligence of School Children. Boston: Houghton Mifflin Company, 1919, 317 p.

A discussion of many phases of intelligence testing including general principles, the amount of individual differences found, capacity according to mental age, prediction according to the I.Q., etc.

153. Terman, L. M. "The use of intelligence tests in the grading of school children," Journal of Educational Research, 1:20-32, January, 1920.

Mostly a theoretical discussion of individual differences, how many pupils should be accelerated and how many retarded, etc. Summarizes what has been found as to difference of test results and teachers' opinions as showing that tests are usually more nearly right.

154. Terman, L. M. "The psychological determinist; or democracy and the I.Q.," Journal of Educational Research, 6:57-62, June 1922.

An answer to Bagley's speech against the use of the I.Q. for classifying pupils.

155. Terman, L. M. et al. Intelligence Tests and School Reorganization. Yonkers, New York: World Book Company, 1922, 111 p.

A theoretical discussion of the problems accompanied by statements of what is being done in Oakland, Los Angeles and Miami, Arizona.

156. Thorndike, E. L. et al. "Intelligence and its measurement: A symposium," Journal of Educational Psychology, 12:123-47, 195-216, 271-75; March, April, May, 1921.

A rather theoretical discussion of the nature and measurement of intelligence and the next steps in research dealing with it.

157. Tildsey, J. L. "Some possibilities arising from the use of intelligence tests," Bulletin of High Points in the Work of the High Schools of New York City, 3:4-10, June, 1921.

Tells of various uses of intelligence tests in New York.

158. Trabue, M. R. "Some pitfalls in the administrative use of intelligence tests," Journal of Educational Research, 6:1-11, June, 1922.

A critical discussion pointing out certain dangers to be avoided.

159. Trabue, M. R. and Stockbridge, F. P. "Psychological tests in education," Chapter VII, Measure Your Mind. New York: Doubleday Page and Company, 1920, p. 63-75.

Among other things contains a brief account of an experiment in Public School No. 64, New York City, wherein mental tests were one of the bases used to select pupils for rapid advancement. Also discusses the predictive value of mental tests in regard to college marks and shows that it is rather high.

160. Tupper, C. R. "The grading of pupils in Miami, Arizona," Educational Research Bulletin, 2:163-65, May 30, 1923.

A description of the plan in use in Miami in which class groupings are made more or less on the basis of mental age. The teachers appear to favor the plan. It has reduced failure by more than two-thirds and increased acceleration considerably.

161. Wallin, J. E. W. "A comparison of three methods for making the initial selection of presumptive mental defectives," School and Society, 13:31-45, January, 1921.

States that the Pressey Primer Scale is inferior to the judgment of trained individuals in selecting defective individuals.

162. Washburne, C. W. "The individual system in Winnetka," Elementary School Journal, 21:52-68, September, 1920.

Describes the system in Winnetka in which progress and promotion are entirely individual, in which most of the work is done by using practise material and then taking complete diagnostic tests. States that every teacher favors the plan, that at least five percent were saved from repeating and $5000 saved in one year.

163. Washburne, C. W. "Educational measurements as a key to individual instruction and promotions," Journal of Educational Research, 5:195-206, March, 1922.

An account of the individual system used in Winnetka.

164. Whipple, G. M. "The problem of selecting and training gifted children in the public schools," Fifth Conference on Educational Measurements, Bulletin of the Extension Division, Vol. 4, No. 4. Bloomington: Indiana University, 1918, p. 6-25.

An argument for the selection of gifted children by mental tests and the providing of special opportunities for them. States that the 10 percent with I.Q.'s above 115 can probably do two years' work in one. Also discusses the Urbana and other experiments.

165. Whipple, G. M. "Educational determinism; a discussion of Professor Bagley's address at Chicago," School and Society, 15:599-602, June 3, 1922.

A strong reply to Bagley's address in which the validity and use of intelligence tests are strongly advocated.

166. Whitcomb, M. Edith. "Intelligence tests in the primary grades," Journal of Educational Research, 5:58-61, January, 1922.

An account of the work done with the Stanford Revision in Council Bluffs, Iowa. Gives various data and shows that the plan of placing pupils by test scores works well.

167. Whitmire, Ethel D. "Intelligence tests vs. teacher's estimates," Psychological Clinic, 13:197-98, May 15, 1920.

Discusses a couple of cases in which teachers rated pupils by their personality and attitude.

168. Whitney, F. P. "Provision for accelerant and retarded children in junior high school," School Review, 27:695-705, December, 1919.

In the Collinswood Junior High School of Cleveland pupils were placed by school marks and test results.

169. Willett, G. W. "A suggestion for meeting individual differences," School Review, 28:576-84, October, 1920.

States that cumulative records are better than either teachers' judgments or intelligence test results, but that a combination of all three is best.

170. WILLIAMS, A. J. "Age-grade distributions and intelligence quotients," Journal of Educational Psychology, 11:39-44, January, 1920.

An account of the Otis Test in Brewster and a discussion of its results as compared with the age-grade location of the pupils.

171. WILLIS, C. B. "The grading and promotion of pupils," Journal of Educational Method, 1:90-95, November, 1921.

An argument for the classification of pupils according to their ability.

172. WINTER, O. "Chicago intelligence test in Harrison Technical High School," School Review, 28:772-75, December, 1920.

It was found that the correlation between intelligence test scores and certain achievement tests and also school marks was rather low.

173. WOODY, CLIFFORD. "Tests and measures in the schoolroom and their value to the teachers," School and Society, 6:61-66, July 21, 1917.

Mentions others but especially tells of an experiment in Chatham, New Jersey. On the basis of scores on five achievement tests the best pupils were given extra promotion. All made good and led in the classes in which they were placed.

174. WOODY, CLIFFORD. "Measurement of the effectiveness of differentiation of high-school pupils on the basis of the Army intelligence tests," Journal of Educational Research, 7:397-409, May, 1923.

A study of the freshman class in a small Michigan high school. Shows that if the class had been divided into three groups according to scores upon Army Alpha the division would have agreed fairly well with achievement as measured by tests in English, Algebra and Latin.

175. WURTH, CARL. "Intelligence tests and promotions," American School Board Journal, 65:50-51, December, 1922.

An argument against segregating bright pupils for three reasons. They should spend their time getting work better rather than going more rapidly, their absence reduces the enthusiasm of the class and they acquire vanity and false pride from being classified as superior.

176. YATES, DOROTHY H. "Impressions from two years spent in testing gifted children." Proceedings of the First Annual Conference on Educational Research and Guidance, San Jose State Teachers' College. Sacramento, California: State Printing Office, 1923, p. 29-33.

A discussion of gifted children with some argument to the point that training for leadership does not need to be begun early and that bright pupils should not be forced too much.

177. Yoakum, C. S. and Yerkes, R. M. Army Mental Tests. New York: Henry Holt and Company, 1920, 303 p.

A general account of how the Army Tests were made and used with a discussion of results and applications.

178. Zirkle, H. W. "Character and results of special rooms as conducted in the Whittier School of Denver," Elementary School Journal, 21:189-97, November, 1920.

Pupils are placed in regular, accelerated and retarded rooms according to test results, age, school marks, health, etc.

179. Zirkle, H. W. "Taking care of the gifted child," The Technique of Supervision by the Elementary School Principal. The First Yearbook of the Department of Elementary School Principals. Washington: National Education Association, May, 1922, p. 81-86.

In a Denver school, superior pupils are selected by individual and group tests, school marks, effort and physical condition and are given the opportunity to gain time. Practically all of the pupils selected are able to gain time and still continue to make high school marks.

180. Zornow, T. A. and Peckstein, L. A. "An experiment in the classification of first-grade children through the use of mental tests," Elementary School Journal, 23: 136-46, October, 1922.

In Rochester, New York, over 400 entering pupils were tested with the Stanford Revision and classified according to the test results. Later the Detroit First-Grade Group Test was used. It was found that but few pupils had to be shifted after being thus placed.

181. "Classification of pupils according to mentality," The School Magazine, 3:196-97, April, 1921.

In a Buffalo school A, B and C groups were formed in grades VI-VIII according to the average of Otis scores, teachers' marks and an average mark. The plan appears to have been successful.

182. "Classification of pupils in Miami, Arizona," Elementary School Journal, 22:727-29, June, 1922.

When pupils were classified on the basis of mental tests, the number of failures was reduced and much money saved.

183. "Double promotion," Elementary School Journal, 23:324-25, January, 1923. An editorial.

In Macomb, Mississippi, all pupils who are recommended by the teacher and have a doctor's certificate of good health, who are not below a certain age, whose previous record in school work and attendance is satisfactory and who make certain scores upon achievement tests are given double promotion if it does not involve skipping certain subjects.

184. Educational Research Bulletin, Ohio State University, Vol. 2, No. 8, p. 123-24, April 18, 1923.

An account of procedure in Painesville, Ohio. In grades VIIB to VIIIA pupils are grouped according to the amount of work they are capable of doing but in the lower grades they are grouped according to ability as shown by the Illinois Examination and school work. Also there is a special room with individual progress and promotion upon effort for upper-grade pupils whose school work is unsatisfactory and whose ability appears to be low.

185. "Grouping pupils on the basis of ability." The First of a Series of Surveys of the Dept. of Instruction, Cleveland Public Schools, Board of Education, Cleveland, Ohio. p. 18-21.

In the first and seventh grades pupils were classified chiefly according to results from the Pressey Primer and Illinois Scales, respectively.

186. "Intelligence tests as a basis for homogeneous grouping at Xenia, Ohio," American School Board Journal, 64:75-76, March, 1922.

Three groups were made up of seventh and eighth grade pupils on the basis of the Haggerty and Terman Tests. They were regrouped later as was necessary but in most cases stayed in the same group.

187. Journal of Educational Research, 7:265, March, 1923.

In Painesville, Ohio, seventh and eighth grade pupils are grouped according to the Illinois Examination results. The sub-normal of these grades are promoted on the basis of their effort.

188. Journal of Educational Research, 7:454, May, 1923.

A brief account of the reclassification of pupils in grades V - VIII on the basis of scores upon the National Group Intelligence Test, teachers' marks and the time devoted to home work. Such changes in the original classification as seemed necessary were made. The plan seems to have been successful.

189. Journal of Educational Research, 7:456-57, May, 1923.

In Miami, Arizona, results from the Otis and National Group Intelligence Tests are used to determine the I.Q.'s of the pupils for the purpose of grouping. Each teacher handles two sections of different ability, the average and the dull. Extra promotions are given to take care of the brighter pupils. The failure rate has been decreased very markedly while results of standardized tests show that achievement has not been lowered.

190. "Provision for exceptional pupils in Oakland," School and Society, 14:30-31, July 16, 1921.

Gives data as to the number of special classes and special promotions for exceptional pupils.

191. "Pupils classified by mental tests," School Life, 12:13, February 1, 1921.

An account of reclassification of pupils in the W. T. Harris school of New York City by group tests.

192. "Reclassified school making good progress," School Life, May 1, 1921. p. 12.

A brief account of the working of the plan in the W. T. Harris school of New York City in which pupils were grouped in sections by group test results. The plan appears to be working satisfactorily.

193. "Report of the school committee for the year 1917-1918." Providence, Rhode Island: School Committee, p. 49-70.

Discusses many points in connection with the use of intelligence tests, especially their high correlation with quality of school work and their reliability for classifying pupils.

194. "Uniform classification by mental age," School Life, 8:88, December, 1921.

A statement that in Denver it is planned to reclassify all pupils into homogeneous groups according to their mental ages.

PART II.

195. ALDERMAN, L. R. "An effort to make the school fit the needs of the exceptional child," Addresses and Proceedings of the National Education Association, 52:830-35, 1914.

An account of provisions for exceptional children in Portland, Oregon, followed by general discussion. Provisions include a special school for defectives and special rooms for other exceptional children.

196. ALEY, R. J. "Care of exceptional children in the grades," Addresses and Proceedings of the National Education Association, 48:881-86, 1910.

A general discussion, its chief point being that superior pupils have been comparatively neglected.

197. APPELL, I. and WOLFSON, A. N. "Plan for organizing the entrants into the high school of commerce according to their attainments in scholarship, their capacities and their aims," School Review, 27:256-61, April, 1919.

In the New York High School of Commerce pupils were formerly grouped according to the foreign language they had studied but the result was not satisfactory. Later they were divided for English work according to their marks in English during the first three weeks, similarly for mathematics. Results were fairly satisfactory.

198. BAGLEY, W. C. "The 'Batavia system' of class-individual instruction," Classroom Management. New York: Macmillan Company, 1916, p. 214-24.

A description of the "Batavia system."

199. Barnard, F. J. "Classification and promotion of pupils." Report of United States Commissioner of Education, Vol. 1, p. 303-56, 1898-99.

An account of a number of flexible plans of promotion including those in operation in St. Louis, Elizabeth, Seattle, Denver, and Cambridge.

200. BARNARD, F. J. "Let pupils be so classified as to allow unrestricted progress or unlimited time, according to ability," Addresses and Proceedings of the National Education Association, 38:163-70, 1899.

An account of the Seattle plan after it had been in operation eight years. Data are given for over 7000 pupils enrolled in four parallel tracks. The plan appears to have been successful.

201. BATES, W. C. "Annual Report of the School Committee and the Superintendent of Schools." Cambridge, Massachusetts, 1908, p. 45-51.

A description of the Cambridge double-track plan. Data are given for seventeen years' operation showing that on the whole the more rapidly pupils progressed through the grades the better marks they made in high school.

202. BECHT, J. G. "Bright pupils and dull pupils," Journal of Education, 79:395-6, April 9, 1914.

A discussion and diagnosis of the causes of retardation and of the fact that bright pupils deserve special help.

203. BELISLE, H. J. "An experiment in retardation," Journal of Education, 75:381, 87-9, April 4, 1912.

An account of the organization of fast, medium and slow classes in Lawrence, Massachusetts. The operation of the plan appears to have been satisfactory.

204. BERGEN, J. Y. "Grading inside of class lines," Educational Review, 16:81-85, June, 1898.

Description of the division of freshmen and other students of the English High School of Boston according to their average marks in various subjects.

205. BERRY, C. S. "Special classes in Michigan for mentally exceptional children." Report of Superintendent of Public Instruction, Michigan, Vol. 78, 1914-15, p. 57-68.

Gives the results obtained from a questionnaire answered by over 200 towns. Five different classes of exceptional children are considered. Fifty-eight percent of school administrators favor classes for bright children.

206. BLEWETT, BEN. "The system of grading pupils in St. Louis," Educational Review, 8:387-8, November, 1894.

An explanation of the St. Louis quarterly system.

207. Boykin, J. C. "Class intervals in city public schools." Report of the United States Commissioner of Education, Vol. 2, p. 981-1009, 1891.

A very good sketch of the history of graded schools from the sixteenth century to the time of writing. The St. Louis and other flexible systems are described.

208. Brehm, J. J. "Variation in the treatment of different groups." Ninth Annual Schoolmen's Week Proceedings, University of Pennsylvania Bulletin, Vol. 23, No. 1, p. 248-52.

In the junior high school of Harrisburg pupils are grouped according to their previous records. The school is "thoroughly committed to this plan."

209. Brown, S. W. "Some experiments in elementary school organization," Addresses and Proceedings of the National Education Association, 51:458-63, 1913.

The chief point made is an argument for individual progress in each subject at the rate best suited to the pupil.

210. Buchanan, Elizabeth. "Classification and promotion," Addresses and Proceedings of the National Education Association, 39:128-37, 1900.

An argument for short class intervals, of not more than twelve weeks, or even six.

211. Burk, C. F. "Promotion of bright and slow children," Educational Review, 19:296-302, March, 1900.

A description of the Santa Barbara system with A, B, and C sections in each grade. These sections differed in the intensity and extensity of work covered.

212. Burk, Frederick. "Every child a minor vs. the lockstep. A suit in equity." Monograph C. San Francisco State Normal School, Fresno, 1915.

A description of individual instruction in the training school for practice teachers. Each pupil goes at his own pace and promotion is made at any time.

213. Burnell, Elizabeth F. "Instruction in mathematics for gifted pupils," Pedagogical Seminary, 24:569-83, December, 1917.

Gives the results of a questionnaire on provisions for bright children. Also a bibliography.

214. Buswell, G. T. "The school treatment of mentally exceptional children." Elementary School Journal, 23:683-93, May, 1923.

A brief discussion of a number of plans of providing for mentally exceptional children. The argument favors modification of the curriculum and varying the methods of work rather than varying the rate of progress.

215. Carfrey, J. H. "Grading and promotions," Journal of Education, 75:379-80, April 4, 1912.

States that shorter intervals than semesters are best. In Franklin, Massachusetts, a modification of the shorter interval and the group plans are used.

216. Chalmers, W. W. et al. "Promotions and gradings," Addresses and Proceedings of the National Education Association, 38:369-75, 1899.

A general discussion of semester vs. annual promotions, various flexible plans of organization, etc.

217. Chalmers, W. S. "Semi-annual promotions." Annual Report of the Board of Education of the City of Toledo, Ohio, for the school year ending August 31, 1903. Toledo, Ohio: Board of Education, 1903, p. 30-41.

An argument showing the importance of semi-annual promotions as a device for getting away from the annual "lockstep." Contains long quotations from W. H. Harris.

218. Clerk, F. E. "The Arlington plan of grouping pupils according to ability in the Arlington High School, Arlington, Massachusetts. School Review, 25:26-48, January, 1917.

A description of the Arlington, Massachusetts plan which relocates pupils every two months on the basis of school marks. Nine advantages are stated and eleven objections answered on the basis of six years of experience with the plan.

219. Cogswell, Francis. "The Cambridge experiment," Addresses and Proceedings of the National Education Association, 33:333-38, 1894.

A description of the Cambridge, Massachusetts, plan which provides for a four-and a six-year track through grades four to nine inclusive.

220. Coleman, E. M. "Grading for efficient organization in the interests of pupils," Addresses and Proceedings of the National Education Association, 40:286-7, 1901.

Describes the plan in use at Ft. Dodge, Iowa, which was adapted from the St. Louis plan.

221. Collicott, J. G. "The bright pupil," Addresses and Proceedings of the National Education Association, 53:457-66, 1915.

Gives results from a questionnaire sent out to all cities of over 50,000. Discusses briefly about a dozen flexible plans for the selection of bright pupils. Followed by arguments for and against.

222. Corson, D. B. "Classification of pupils," Educational Administration and Supervision, 6:86-93, February, 1920.

An argument for rapid, regular and slow sections of pupils.

223. Cubberley, E. P. "Promotional plans." Public School Administration. Boston: Houghton Mifflin Company, 1916, p. 300-06.

Describes the Pueblo, Batavia, Cambridge, Mannheim, and a few other plans.

224. Cummins, R. A. "'Bright' and 'slow' pupils in elementary and high school," Journal of Educational Psychology, 10:377-88, October, 1919.

An account of a study in a private school in New York City. It was found that there was a tendency for pupils who were bright and making fast progress to continue to do so and for others to continue to make slow progress.

225. Dempsey, C. H. "Flexible grading and promotions," Journal of Education, 75:373-5, April 4, 1912.

Describes the reorganization of the system at Malden, Massachusetts, which provided opportunity for bright pupils to skip three semesters of work. Lists the principles upon which reorganization was based.

226. De Voss, J. C. "School success of gifted children." Proceedings of the First Annual Conference on Educational Research and Guidance, San Jose State Teachers College. Sacramento, California: State Printing Office, 1923, p. 34-38.

An argument to the effect that gifted children are able to carry extra work and progress more rapidly. To some extent replies to Dr. Bagley's address on educational determinism.

227. Dougherty, N. C. "Report of the round-table discussion on 'promotion in city schools'," Addresses and Proceedings of the National Education Association, 31:802-3, 1892.

States that in Chicago promotions are made at all times, only slightly over half being at the end of the year. Argues that a flexible system helps pupils.

228. Downes, F. E. "Seven years with unusually gifted pupils," Psychological Clinic, 6:13-17, March 15, 1912.

Describes the bright sections organized in the Harrisburg, Pennsylvania, schools, also the special schools for bright pupils of the upper grades. States that attendance, scholarship, discipline and high-school records were all improved by the operation of the plans described.

229. Downes, F. E. "Other special schools." Report of the Superintendent of Schools, Harrisburg Public Schools, 1912, p. 23-26.

Describes the schools for exceptionally gifted children which allow them to cover work of the eighth and ninth years in one year. States advantages of this plan and that the pupils involved do well in high school.

230. Dutton, S. T. and Snedden, David. "Grading and promotion." The Administration of Public Education in the United States, New York: Macmillan Company, 1912, p. 341-55.

A discussion of homogeneity within a grade, reclassification and flexible grading, minimum and maximum requirements, individual teaching, ungraded rooms, the Cambridge and Batavia plans, etc.

231. Dyer, F. B. "Superintendent's report." Seventy-ninth Annual Report of the Public Schools of Cincinnati, 1908, p. 66, 67-69.

A brief account of certain special classes and vacation schools in Cincinnati.

232. Dyer, F. B. "Rapid advancement classes." Annual Report of the Superintendent, Boston, Public Schools, 1913, p. 65-67.

An account of certain classes in which pupils may do three years of work in two. The plan worked well and was to be extended to other schools of the city.

233. Fisher, G. M. "Report to superintendent." Sixty-fourth Annual Report of Schools of Worcester, 1912, p. 74-75.

An account of the summer school provided for failing pupils to make up work and for bright pupils to gain time.

234. Freeman, F. N. "Provision in the elementary school for superior children," Elementary School Journal, 21:117-31, October, 1920.

Gives results of a questionnaire sent out to cities of over 25,000. A rather careful analysis of results is made showing what plans are in use and upon what they are based. Advantages and difficulties are discussed.

235. Garber, J. P. "A rational system of classification and promotion of pupils in elementary schools," Education, 27:288-302, January, 1907.

Argues that promotional intervals should fit pupils but that bright and dull pupils should not be separated nor should intervals be very short. Also says that failure is not the chief cause of elimination.

236. Gillingham, Anna. "The bright child and the school," Journal of Educational Psychology, 10:237-52, May-June, 1919.

A description of a number of bright children and an argument that such children be given more recreation and play rather than extra work leading to accelerated progress.

237. Gist, A. S. "The acceleration of pupils," School and Society, 5:116-18, January 27, 1917.

An account of an experiment in grouping pupils according to teachers' estimates in a Seattle school. The experiment was considered a success.

238. Glass, J. M. "The study-coach or opportunity class organization." Eighth Annual Schoolmen's Week Proceedings, University of Pennsylvania Bulletin, Vol. 21, No. 37, Philadelphia, June 18, 1921, p. 250-56.

Describes two types of study-coach organization to meet the individual needs of three classes of pupils. Gives a list of advantages of so doing.

239. Gordon, C. H. "Reorganization of the grammar school and a rational system of grading," Education, 21:16-26, September, 1900.

An argument for two parallel courses of the same length, one of which has a richer curriculum than the other, and for the basing of promotion chiefly upon effort.

240. Gosling, T. W. "The classification of pupils," Educational Review, 39:394-399, April, 1910.

A discussion of the problem as it exists in high school.

241. Greenwood, J. M. "Shorter time in elementary school work," Educational Review, 24:375-90, November, 1902.

Gives data concerning the Kansas City elementary schools which do in seven years what most schools do in eight. Shows that the work is satisfactorily done.

242. Gregory, Christopher. "Holding pupils in school," Atlantic Educational Journal, 11:436-37, April, 1916.

Describes the semi-annual plan with special classes for retarded and rapid classes for superior pupils in Long Branch, New Jersey. States that it results in the seventh and eighth grades being as large as the fifth and sixth.

243. Harley, H. L. "The physical status of the special class for bright children in the University of Pennsylvania, summer session of 1912," Psychological Clinic, 7:20-23, March 15, 1913.

Shows that school work done in hot weather did not injure the health of bright pupils. All except one increased in weight while in summer school.

244. Harris, W. T. "Advantages and disadvantages of the graded system." Fifteenth Annual Report of the St. Louis Public Schools, 1868-69, p. 104-07.

This discussion implies dissatisfaction with the rigid grading system.

245. Harris, W. T. "Superintendent's report." Eighteenth Annual Report of the St. Louis Public Schools, 1871-72, p. 24-27, 81-87.

Mentions the quarterly promotions in use in St. Louis, and discusses the advantages of classification if not too rigid. States that a five-weeks interval is ideal and that a system must be readjusted to meet the needs of its pupils.

246. Harris, W. T. "Superintendent's report." Nineteenth Annual Report of the St. Louis Public Schools, 1872-73, p. 24-29.

An explanation of the flexible system used in St. Louis which provided for regular quarterly promotions but also had some groups organized with only five weeks intervals.

247. Harris, W. T. "Superintendent's report." Twentieth Annual Report of the St. Louis Public Schools, 1873-74, p. 121-48.

An argument in favor of the St. Louis and other flexible plans of organization. Contains quotations from several others in favor of such plans.

248. Harris, W. T. "Superintendent's report." Twenty-first Annual Report of the St. Louis Public Schools, 1874-75, p. 28-29, 86-93.

Contains quotations from other exponents of flexibility.

249. Harris, W. T. "Classification in graded schools." Report of the United States Commissioner of Education, Vol 1, p. 601-36, 1891-92.

This is merely a collection of what is contained in the various reports of the St. Louis Public Schools on the same subject.

250. Harris, W. T. "Class intervals in graded schools," Addresses and Proceedings of the National Education Association, 39:332-40, 1900.

A statement of the need of flexible organization and of separating weak and strong pupils. It is also argued that promotion should be individual and not by classes.

251. Harris, W. T. et al. "The early withdrawal of pupils from school: its causes and its remedies," Addresses and Proceedings of the National Education Association, 12:260-73, 1872.

Several causes of elimination are stated of which defective grading is the most potent. Discussion both for and against short intervals of organization follows.

252. Hartwell, C. S. "Economy in education," Educational Review, 30:159-77, September, 1905.

States that the grammar school is not satisfactory and that promotion by subjects is one of the changes needed.

253. Hartwell, C. S. "The grading and promotion of pupils," Educational Review, 40:375-86, November, 1910. Addresses and Proceedings of the National Education Association, 48:294-306, 1910.

An argument for flexible organization and promotion by subjects. The results of a questionnaire are given showing more favorable sentiment on these points than existed several years previously. It is also stated that bright and dull pupils should be given individual attention.

254. Hatch, W. E. et al. "Provisions for exceptional children in the public schools," Addresses and Proceedings of the National Education Association, 45:360-63, 1907.

Rather brief discussion of several of the plans.

255. Heilman, J. D. "The need for special classes in the public schools," Psychological Clinic, 1:104-14, June 15, 1907.

This deals entirely with classes for inferior pupils.

256. Henry, T. S. "Classroom problems in the education of gifted children." Nineteenth Yearbook of the National Society for the Study of Education, Part II, 1920. 125 p.

A complete account of an experiment in selecting gifted children at Urbana, Ill., and a brief description of a number of flexible promotion schemes. There is also a bibliography on the psychology and pedagogy of gifted children.

257. Hoblit, M. L. "The high-school unit: quantity, quality, and credit," School Review, 23:303-06, May, 1915.

A suggestion that high-school graduation be determined by a combined system of units and points, the latter to depend upon the marks made. The scheme is so arranged that a superior student can complete the work in three years, an average student in four years and a poor student in more than four.

258. Hoblit, M. L. "Is credit for quality sound?—A rejoinder," School Review, 23:712-14, December, 1915.

States that his suggested plan did not involve the graduation of a pupil who had covered less than fifteen units of work. Also argues in its defense.

259. Holmes, W. H. "Plans of classification in the public schools," Pedagogical Seminary, 18:475-522, December, 1911.

Describes a dozen or more of the different plans used in this country, also several used in England. The discussion is followed by a bibliography.

260. Holmes, W. H. "The St. Louis Plan," Journal of Education, 75:380, April 4, 1912.

A brief description of the plan mentioned.

261. Holmes, W. H. "School organization and the individual child." Worcester: Davis Press, 1912, Partly in Journal of Education, 75:376-9, April 4, 1912.

A long discussion of a number of different plans of classification and promotion. Also a theoretical discussion of the subject. Some space is given to mental tests.

262. Hutton, T. B. "Classification and gradation," School Review, 16:543-50, October, 1908.

Describes the two-group system in use at LeMars, Iowa. States that it prevents retardation and saves money.

263. Jackman, W. S. "The school grade a fiction," Educational Review, 15:456-73, May, 1898.

A strong plea and argument against having eight rigidly defined elementary grades.

264. Johnson, F. W. "Credit for courses in the University High School," School Review, 23:715-18, December, 1915

Gives the plan of credit for quality followed in the University of Chicago High School. The credit ranges from .85 for a grade of 60 percent to 1.25 for a grade of 95 percent. States that the University of Chicago admits some students on this basis.

265. Jones, E. E. "A concrete example of the value of individual teaching," Psychological Clinic, 2:195-203, December 15, 1908.

Gives the history of one successful case and states a few general principles.

266. **Jones**, E. E. "Suggestions from cases of unusually rapid or irregular progress in public schools," Addresses and Proceedings of the National Education Association, 50:640-45, 1912.

A study of several hundred pupils who had skipped grades which shows that most of them should have skipped and that a few should not.

267. **Jones**, H. S. et al. "II—Report of the committee on city school systems. Pupils-classification, examination and promotion." Addresses and Proceedings of the National Education Association, 26:276-84, 1886.

In this discussion are mentioned various types of classification and the need for considering individual pupils.

268. **Jones, Olive** M. Teaching Children to Study. The Group System Applied. New York: Macmillan Company, 1909.

Gives a brief account of a number of plans and a very full one of the Mannheim special class system.

269. **Jones**, W. F. "An experimental-critical study of the problem of grading and promotion," Psychological Clinic, 5:63-96, 99-120, May 15, and June 15, 1911.

A discussion of a number of double-track systems and of the reasons for failure. States that the pupils' and teachers' reasons for the latter are as often false as true.

270. **Kendall**, C. N. "What modifications in organization are necessary to secure suitable recognition for pupils of varying ability, particularly for the ablest?" Addresses and Proceedings of the National Education Association, 46:147-55, 1908.

Discusses the Cambridge and other plans. Is in favor of making some modifications.

271. **Kennedy, John.** "The Batavia plan after fourteen years of trial," Elementary School Teacher, 12:449-59, June, 1912.

After fourteen years experience with the Batavia plan the author is still an enthusiastic believer in its advantages.

272. **Kennedy, John** et al. The Batavia System of Individual Instruction. Syracuse: C. W. Bardeen, 1914. 299 p.

A complete description of the Batavia system, with two teachers in each room, one the regular teacher and the other to aid individuals. States that elimination and expense were reduced, teachers freed from worry and the general school spirit bettered.

273. **Kennedy**, J. W. "The all-year school," Addresses and Proceedings of the National Education Association, 55:795-801, 1917.

Tells of summer term and "promotion classes" in Newark. Both time and money were saved.

274. Kilpatrick, V. E. "Emancipating the individual pupil," Education, 30:375-85, February, 1910.

A general discussion of individuality followed by an enumeration of ten plans of school organization.

275. Kirk, J. R. "Should the school furnish better training for the non-average child?" Addresses and Proceedings of the National Education Association, 45:221-27, 1907.

A plea for flexible organization.

276. Leighton, E. V. "At last—a chance for the gifted child," Popular Educator, 30:115, October, 1912.

Brief mention of several cities making provision for or giving recognition to gifted children.

277. Lewis, H. P. "Semi-annual promotion." Report of the Public Schools of the City of Worcester. Worcester, Massachusetts: School Committee, 1904, p. 34-41.

An argument in favor of semi-annual promotions. Contains quotations from several sources.

278. Littwin, M. F. "English 'star' classes at the Boys' High School." Bulletin of High Points in the Work of the High Schools of New York City, 5:4-7, March, 1923.

On the basis of school marks special sections were formed in the sixth, seventh and eighth semesters of high school. These sections were given differentiated and enriched work in English. A number of advantages were derived from the plan.

279. Lynch, Ella F. "The bright child," Psychological Clinic, 4:141-44, October 15, 1910.

A plea that the bright child be given individual instruction and the opportunity for individual progress.

280. McDonald, R. A. F. "Provision for the exceptionally gifted." Adjustment of School Organization to Various Population Groups. Teachers College Contributions to Education, No. 75. New York: Teachers College Bureau of Publications, p. 90-101.

A theoretical discussion followed by a brief description of about a dozen plans.

281. Maennel, B. "The auxiliary schools of Germany." (Translated by Dresslar, F. B.). United States Bureau of Education Bulletin, No. 3, 1907, p. 43-47, 121-23.

An account of the Mannheim three-fold plan.

282. Mangold, G. B. "The mental classification of children. The education of exceptional children." Problems of Child Welfare. New York: Macmillan Company, 1914, p. 205-27.

This is almost entirely a discussion of backward children.

283. Meyer, M. F. "Is credit for quality sound?—A criticism," School Review, 23:708-11, December, 1915.

An argument in favor of credit for quality which is in use in a portion of the University of Missouri. Especially answers Professor Ruediger's adverse criticism.

284. Miller, C. A. A. J. "The study of exceptional children," Addresses and Proceednigs of the National Education Association, 46:957-63, 1908.

A theoretical discussion listing ten kinds of exceptional children.

285. Mitchell, David. "Schools and classes for exceptional children." Educational Survey. Cleveland, Ohio: Cleveland Foundation, 1916. 122 p.

Discusses provisions for ten classes of exceptional children in Cleveland. Bright children are not included in those provided for.

286. Mulrey, Cora L. "The rapid advancement class," Educational Administration and Supervision, 3:416-19, September, 1917

A brief account of the Cambridge plan by which three years of work may be done in two.

287. Myers, G. C. "Broadening the course of study for the bright child," Educational Administration and Supervision, 3:33-37, January, 1917.

Suggests minimum and maximum courses and answers a number of objections to them.

288. Neverman, P. F. "New Richmond plan of grade promotion," American School Board Journal, 54:38, January, 1917.

Describes the New Richmond, Wisconsin, plan which provided two tracks through the lower grades. Also a number of principles upon which the plan is based are given.

289. Parlin, F. E. "Gradation and promotion." Annual Report of the School Committee and the Superintendent of Schools, Cambridge, Massachusetts, 1910, p. 19-21.

A brief explanation of the Cambridge plan.

290. Parkinson, W. D. "Promotions, accelerated and retarded," Education, 19:152-57, November, 1898.

An argument that there is some advantage in the inertia of the graded system and that only a comparatively few pupils should be accelerated and retarded.

291. Parkinson, W. D. "Individuality and social adjustment as means and ends in education," Education, 29:16-24, 104-12, September, October, 1908.

A discussion of individual and class instruction and progress. Argues for a fairly flexible organization adapted to the individual.

292. Patterson, M. Rose. "A preparatory center in Baltimore," Atlantic Educational Journal, 12:234-38, January, 1917.

A description of the Baltimore plan in which the best upper grade pupils of the city are brought together at the end of the sixth grade and may either do three years of work in two or two in one and one-half.

293. Patterson, M. Rose. "The Homewood demonstration school at Johns Hopkins University," School and Society, 16:577-84, November 18, 1922.

An account of an experiment in which a number of pupils did five months work in two. Especially the under-age pupils succeeded in doing the extra work.

294. Payne, W. H. "Elastic grading." Report of the United States Commissioner of Education, Vol. 2, p. 1376-77, 1899-1900.

Argues that the best basis of classification is the average of the year's work and that on such a basis few pupils should need reclassification.

295. Phillips, D. E. "The child vs. promotional machinery," Addresses and Proceedings of the National Education Association, 50:349-55, 1912.

A discussion of changes in school organization that affect promotion.

296. Pickard, J. L. "Superintendent's report." Twenty-first Annual Report, Department of Public Instruction, Chicago, 1874-75, p. 45-57.

Defends the "graded system" but admits that its evils need correction. Believes that the chief correction should be the promotion of pupils whenever they are ready.

297. Prince, J. T. "The grading and promoting of pupils," Educational Review, 15:231-45, March, 1898.

A general discussion with a brief description of several flexible plans.

298. Prince, J. T. "Some New England plans and conclusions drawn from a study of grading and promotion," Addresses and Proceedings of the National Education Association, 37:423-34, 1898.

Gives results obtained from a questionnaire, also a description of the Cambridge, Middleboro, Woburn, Keene, St. Louis, Dayton, Le Mars and Centralia plans.

299. Prince, J. T. "Classification and promotion of pupils." School Administration. Syracuse: C. W. Bardeen, 1906, p. 125-33.

A discussion of several different flexible plans of organization including frequent promotions, the "double-tillage" plan, the organization of groups according to ability and individual promotion.

300. Pyle, W. H. "A psychological study of bright and dull pupils," Journal of Educational Psychology, 6:151-56, March, 1915.

Shows that mental differences between pupils may be measured.

301. Pyrtle, E. Ruth. "Super-normal children—a study," Journal of Educational Method, 1:273-75, March, 1922.

Describes the plan used in the junior high school of Lincoln, Nebraska. On the basis of their school records the brightest pupils are selected and allowed to do three years of work in two. Many advantages are listed.

302. Rathmann, C. G. "The Mannheim system of school organization," Educational Review, 53:55-60, January, 1917.

A description of this plan with an account of its effect upon elimination which was much reduced by its operation.

303. Richman, Julia. "A successful experiment in promoting pupils," Educational Review, 18:23-29, June, 1899.

In a New York school the pupils of grades I-VI were divided into bright, medium and poor sections, each going at its own rate. The effect was to increase the number of promotions.

304. Rigler, Frank. "Principles of classification." Thirty-seventh Annual Report of the Public Schools of Portland, Oregon, 1909-10, p. 113-19.

Description of the Portland plan in which the course of study is divided into fifty-four parts arranged in six cycles of one and one-half years each. At the end of each cycle pupils are reclassified according to their ability, there being two rates at which they may progress.

305. Roberts, J. E. "A working scheme of promotional efficiency," Elementary School Journal, 17:719-26, June, 1917.

Tells of the plan used in Fond du Lac, Wisconsin, where pupils are promoted at any time by subjects and not by grades.

306. Rosier, Joseph. "The graded school—its strength and its weakness," Review of Education, 7:175-77, December, 1901.

States that the lack of individual instruction is not a weakness inherent but one that can be remedied by teachers.

307. Rosier, Joseph. "A satisfactory basis for the promotion of pupils," School and Society, 1:701-05, May 15, 1915. Addresses and Proceedings of the National Education Association, 53:477-84, 1915.

A theoretical discussion of flexible grading and promotion, arguing for special schools, ungraded rooms, maximum and minimum curricula, etc.

308. Reudiger, W. C. "Is credit for quality sound?" School Review, 23:450-54, September, 1915.

An argument against giving credit for quality. States that quality and quantity are disparate, that we recognize quality by the honor attached, that credit for quality would over-emphasize marks, and that it would vary the educational content covered by different students but indicated by the same diploma.

309. Sakaki, Yasusaburo. "Some studies on so-called 'ab-normally intelligent' pupils," Psychological Clinic, 6:18-25, March 15, 1912.

Classifies "ab-normally intelligent" pupils into seven classes and shows the need for individual instruction.

310. Search, P. W. "Individual teaching: the Pueblo plan," Educational Review, 7:154-70, February, 1894.

A description of the Pueblo plan in which all work, promotion and graduation is strictly individual. Lists a number of advantages.

311. Search, P. W. "The Pueblo plan of individual teaching," Educational Review, 8:84-85, June, 1894.

A brief explanation of the Pueblo plan supplementary to the reference above.

312. Search, P. W. et al. "Individualism in mass education," Addresses and Proceedings of the National Education Association, 34:398-411, 1895.

A rather general criticism of the graded system followed by discussion of both sides.

313. Shaer, I. "Special classes for bright children in an English elementary school," Journal of Educational Psychology, 4:209-22, April, 1913.

Describes a plan used in a Manchester school in which there are four special rooms by means of which pupils may gain one or more years' time.

314. Shearer, W. J. "The Elizabeth plan of grading," Addresses and Proceedings of the National Education Association, 37:441-48, 1898.

In Elizabeth, New Jersey, promotions were made at any time upon the basis of the teacher's estimate and pupils grouped according to ability, each group going as fast as it was able. Many favorable results were obtained from this plan.

315. Shearer, W. J. The Grading of Schools. New York: H. P. Smith, 1899. 220 p.

Suggests that grading should be in groups according to ability and acquirements.

316. Shearer, W. J. "Grading for efficient organization in the interests of pupils," Addresses and Proceedings of the National Education Association, 40:285-86, 1901.

A general discussion in which the essentials of good grading are listed as follows: acceleration, classification, proper provision for reclassification and proper apportionment of work.

317. Shearer, W. J. "School children in lockstep," World's Work, 14:9252-55, August, 1907.

A strong argument against the lockstep with a discussion of conditions in New York, Philadelphia, etc.

318. Siders, W. R. "In class instruction, how can the individual be reached?" Addresses and Proceedings of the National Education Association, 47:175-82, 1909.

Discusses several other plans and describes that in use in Pocatello, Idaho. A combined class, group and individual instruction system is used there.

319. Smith, H. L. "Plans for saving time in grades VII-XII inclusive." Bulletin of the Extension Division, Indiana University, Vol. 4, No. 4 (Fifth Conference on Educational Measurements), 1918, p. 74-91.

Presents the results of a questionnaire study of the plans in use in over 400 cities.

320. Snedden, David. "Exceptional children." Monroe's Cyclopedia of Education, New York: Macmillan Company, 1911, Vol. 2, p. 540-41.

Says that flexible promotion more or less takes care of exceptional children.

321. Snedden, David. "Grading and promotion." Monroe's Cyclopedia of Education, New York: Macmillan Company, 1912, Vol. 3, p. 126-28.

Mentions several flexible plans and discusses the subject in general.

322. Spain, C. L. "The platoon school in Detroit." Detroit Educational Bulletin No. 2, November, 1920. Journal of Educational Research, 3:76-77. January, 1921.

Describes the platoon organization in use in some Detroit schools.

323. Spaulding, F. E. "The unassigned teacher in the schools," School Review, 15:201-16, March, 1907.

In Newton, Massachusetts, unassigned teachers are assigned to work with individuals and small groups. They help the better pupils to gain time and the poorer ones to keep up with the class.

324. Spaulding, F. E. "The Newton educational policy." Report of Superintendent. Annual Report of the School Committee in Newton, Vol. 74, p. 15-22, 26, 1913.

States that Newton aims to provide educational opportunities adapted to all individual pupils. Parts of the general plan are to have assistant teachers to help individuals and to make promotions at any time.

325. Stern, Wm. "The supernormal child," Journal of Educational Psychology, 2:143-48, 181-90, March and April, 1911.

An argument to the effect that special types of instruction must be provided for the super-normal and sub-normal pupils.

326. Stevenson, R. W. "Superintendent's report." Columbus Public Schools, 1872-73, p. 45.

States that Columbus, Ohio, allows promotions at any time.

327. **Terman, L. M.** "The mental hygiene of exceptional children," Pedagogical Seminary, 22:529-37, December, 1915.

States that unusually bright children are not given enough work to do and dull children too much.

328. **Terman, L. M.** "Discussion of gifted children section." Proceedings of the First Annual Conference on Educational Research and Guidance, San Jose State Teachers College. Sacramento, California: State Printing Office, 1923, p. 41-43.

A summary of several addresses dealing with gifted children, with some argument in favor of rapid progress rather than enriching the curriculum in the ordinary way.

329. **Trummer, Mary C.** "Instruction adapted to groups of differing ability," School Review, 30:409-11, June, 1922.

Describes the "accomodation" and "express" classes, the "opportunity," "adjustment" and "make-up" groups used in a Los Angeles high school. Pupils are selected on the basis of their elementary school marks.

330. **Ulrich, Flora.** "A year's work in a 'superior' class," Psychological Clinic, 5:245-50, January 15, 1912.

In a Cincinnati school a group of superior children did two years of work in one year but this gain was the least of the benefits derived from the plan. Mental activity and development were greatly increased.

331. **Van Sickle, J. H.** "Plan of the North-Side schools of Denver," Addresses and Proceedings of the National Education Association, 37:434-41, 1898.

A description of the Denver plan and a general discussion in favor of flexibility, minimum essentials, etc. In the Denver plan there were no fixed promotion dates until the third grade was reached. Above that the promotions were made every semester.

332. **Van Sickle, J. H.** "Provision for exceptional children in the public schools," Psychological Clinic, 2:102-11, June 15, 1908.

A discussion of provisions for all sorts of exceptional children except the superior.

333. **Van Sickle, J. H.** et al. "Preliminary report of the committee on provision for exceptional children in the public schools," Addresses and Proceedings of the National Education Association, 46:348-85, 47:343-67, 1908, 1909.

A discussion of physically, intellectually, and morally exceptional children and of various provisions being made for them in Worcester, Indianapolis, Baltimore, St. Louis, Berlin, and elsewhere.

334. Van Sickle, J. H. "Provision for gifted children in public schools," Addresses and Proceedings of the National Education Association, 48:155-60, 1910.

Mentions various reports and studies that had been made and tells of the Baltimore plan of giving high-school pupils extra work if they are bright enough.

335. Van Sickle, J. H. "Provision for gifted children in public schools," Elementary School Teacher, 10:357-66, April, 1910.

States that both inferior and superior children should receive special attention and that there should be groups organized on the basis of ability.

336. Van Sickle, J. H., Witmer, L. and Ayres, L. P. "Provision for exceptional children in public schools." United States Bureau of Education Bulletin, 1911, No. 14. 92 p.

A rather complete discussion dealing with methods of finding, classifying, promoting, teaching, and in general providing for exceptional children. Also deals with the number found retarded, eliminated, etc.

337. Wallin, J. E. W. "The mental health of the school child." Public School Provisions for Mentally Unusual Children. New Haven: Yale University Press, 1914, p. 383-428.

Gives questionnaire results from over 300 cities and sketches the history of special and ungraded classes.

338. Wanger, Ruth. "The special class." News Letter, 19:4-5, May, 1923.

An account of provisions for dull pupils of high-school age. A special course including English, civics, commercial work, sewing, etc. was arranged for those who could not do ordinary academic work. The results have on the whole been favorable.

339. Washburne, C. W. "Breaking the lockstep in our school," School and Society, 8:391-402, October 5, 1918.

An argument for individual progress.

340. Whipple, G. M. "Supernormal children." Monroe's Cyclopedia of Education. New York: Macmillan Company, 1913, Vol. 5, p. 464-68.

A discussion of a number of plans for taking care of bright children.

341. White, E. E. et al. "Several problems in graded-school management," Addresses and Proceedings of the National Education Association, 14:254-63, 1874.

A discussion of annual promotion and the need of adapting the schools to individual pupils.

342. Witmer, Lightner. "The training of very bright pupils," Psychological Clinic, 13:88-96, December 15, 1919.

A discussion of the meaning of brightness, the number of bright and inferior pupils found, etc.

UNIVERSITY
of ILLINOIS
PRESS

TABLE OF CONTENTS

PAGE

PREFACE 5

CHAPTER I.—INTRODUCTION 7

CHAPTER II.—SUMMARY OF CRITICISMS OF WRITTEN EXAMINATIONS 9

Examinations yield inaccurate measures of school achievement 9

Written examinations tend to encourage undesirable mental processes 13

Passing the final examination an undesirable objective 14

Examinations injurious to health of students 14

Time devoted by teachers to written examinations not profitably spent 15

CHAPTER III.—PREPARATION AND ADMINISTRATION OF EXAMINATIONS IN HIGH SCHOOL 16

The data collected 16

Requirement of final examinations in Illinois high schools 16

Time devoted to written examinations 19

Characteristics noted in marking examination papers 20

Weighting of questions 22

Recognition of rate of work 22

Methods of marking examination papers 23

Directions to students concerning methods of work 24

Recognition of a standard distribution in assigning grades to examination papers 24

Relation of examination grades to final grades 25

Summary 25

CHAPTER IV.—THE CONSTANT AND VARIABLE ERRORS IN EXAMINATION GRADES.......... 27
Constant and variable errors of measurement............ 27
Magnitude of variable errors of measurement in standardized test scores and in examination grades.......... 28
Methods employed in present investigation concerning reliability of written examinations.................. 29
Data collected for investigation......................... 30
Reliability of written examination grades and of standardized test scores........................... 31
Conditions tending to produce variable errors of measurement in examination grades..................... 37
Magnitude of constant errors of measurements in standardized test scores and in examination grades.......... 40
Results of present investigation and of previous studies compared.......................... 41
Conclusion—relative accuracy of examination grades and of test scores.......................... 41

CHAPTER V.—THE CONTENT OF WRITTEN EXAMINATIONS..... 43
The data collected.............................. 43
Classification of questions.......................... 43
Relation of examination questions to educational objectives............................ 46

CHAPTER VI.—THE IMPROVEMENT OF WRITTEN EXAMINATIONS. 48
Reduction of constant errors......................... 50
Reduction of variable errors......................... 54
Agreement of content of examinations with educational objectives.......................... 55
Simplification of administration of written examinations.. 56

CHAPTER VII.—RULES FOR THE PREPARATION AND ADMINISTRATION OF WRITTEN EXAMINATIONS................. 63

APPENDIX.................................. 66

PREFACE

This bulletin reports the results of three extensive investigations relating to written examinations. These investigations were made by Mr. Souders under the direction of the Director of the Bureau of Educational Research. The tabulations and statistical calculations were made by Mr. Souders or by clerks working under his immediate direction. The preparation of the published report, however, is the work of the Director of the Bureau.

The Bureau of Educational Research wishes to acknowledge its indebtedness to the superintendents, principals, and teachers who cooperated by furnishing the necessary data. The data required in the study of reliability of written examinations necessitated considerable additional labor. Without their cooperation these investigations would not have been possible.

WALTER S. MONROE, *Director.*

November 1, 1923.

PRESENT STATUS OF WRITTEN EXAMINATIONS AND SUGGESTIONS FOR THEIR IMPROVEMENT

CHAPTER I

INTRODUCTION

Preparation and administration of written examinations important phases of the teacher's work. Written examinations, except in the few schools where they have been abolished, form a very important phase of the teacher's work, both because of the time devoted to their preparation and administration and of the significance attached to the measures which they yield. The final grades upon which promotion and the awarding of school honors depend are determined largely by final examinations and by written tests given during the school term. Altho standardized educational tests have become widely used during recent years, written examinations are still the most frequently used type of measuring instrument. This will probably always be true, particularly in the high school. Hence, we may expect that written examinations will occupy in the future as in the past, an important place in the work of our schools.

Need for more information concerning written examinations. There have been numerous investigations which showed that the marking of written examination papers is highly subjective—that is, different teachers tend to assign different marks to the same paper. With the exception of these studies relatively little precise information is available in regard to written examinations but a number of criticisms based upon experience and theoretical considerations have been made. As a result many teachers and other school officials have come to consider written examinations very inferior instruments and have abolished them in a number of schools.

A search through our educational literature, particularly textbooks, reveals an astonishing lack of information in regard to written examinations. Relatively little specific attention has been given to their preparation and administration in our courses for the training of teachers. Inexperienced teachers have been left largely to their own devices in this important phase of their work. Careful inquiry and observation have indicated that there is a variety of practises with reference to the types of questions asked and to the administration of written examinations. Hence it appears that there is need for a comprehensive investigation of the present status of written examinations in order that a more intelligent estimate may be formed of their value in the process of education and that specific directions may be formulated in regard to their preparation and administration.

Purpose of this bulletin. It is the purpose of this bulletin to present (1) a brief summary of certain previous investigations relating to written examinations and also of the arguments for and against written examinations; (2) the results of three extensive investigations conducted by the Bureau of Educational Research, (a) the preparation and administration of written examinations in Illinois high schools, (b) the constant and variable errors in examination grades, and (c) the content of written examinations; and (3) suggestions for the improvement of written examinations. In the concluding chapter the author presents a list of rules in regard to the preparation and administration of written examinations.

CHAPTER II

SUMMARY OF CRITICISMS OF WRITTEN EXAMINATIONS[1]

Plan of chapter. In this chapter the important criticisms of written examinations are briefly summarized. Following each criticism either a brief answer is given or a reference is made to a detailed discussion in a later chapter. By presenting both sides of the question in this way, it is hoped that the reader will be assisted in forming an intelligent estimate of the merits of written examinations.

I. Examinations yield inaccurate measures of school achievement. In support of this argument six points have been made.

1. The most important criticism relating to the accuracy of written examinations is that the marking of the papers is highly subjective. A large number of scientific investigations have yielded objective evidence that different teachers when working independently tend to assign widely varying marks to the same paper. One of the first studies of this type was by Starch and Elliott who found that the marks assigned to the same examination paper in Plane Geometry by 116 teachers ranged from 28 to 92 on the scale of 100 percent. The facts of such investigations as this can not be disputed but as we have no means of securing perfectly accurate measures of achievement, the question at issue concerns the relative rather than the absolute accuracy of the measurements secured. Facts may be misinterpreted. In Chapter IV we shall present evidence to show that when judged in relation to other means for measuring school achievement, written exami-

---

[1] Starch, Daniel, and Elliott, E. C. "Reliability of grading high-school work in mathematics," School Review, 21:254-59, April, 1913.

Morton, Robert L. "The examination method of licensing teachers," Educational Administration and Supervision, 6:421, November, 1920.

Wood, Ben D. "Measurement of college work," Educational Administration and Supervision, 7:301-34, September, 1921.

Kelly, F. J. "Teachers' marking," Teachers College Contributions to Education, No. 66. New York: Teachers College, Columbia University, 1914.

nations yield relatively more accurate measures than generally supposed. In view of the additional information secured the subjectivity of written examinations loses much of its potency as a reason for their abolition.

2. The questions of ordinary examinations are usually not equal in difficulty and weighting by teachers is highly subjective.[2] It has been inferred that this condition tends to increase materially the inaccuracy of examination marks. Comparisons of weighted and non-weighted scores yielded by standardized tests have revealed that the errors introduced by disregarding the unequal difficulty of exercises or questions are not significant in most cases.[3]

3. It has been pointed out that frequently the content of written examinations is not in agreement with recognized educational objectives. Catch questions relating to trivial facts or worded in a misleading way have been cited as illustrations. Certain examination questions also have referred to items which had not been included in the course or at least had received only minor emphasis. Some evidence with reference to the justification of this criticism will be presented in Chapter V.

4. In most examinations the rate of work is neglected. The usual practise is to allow sufficient time for all pupils to finish or to base the mark only on the questions answered in the unfinished papers. Hence a student's examination grade is not influenced by the rate at which he answers the questions. It is easily possible to take into account the student's rate of work in determining the mark assigned to his examination paper. One plan is to set an examination of sufficient length so that all members of the class will be employed during the entire period. Another procedure is to have the student record the time when he finishes. In this way some weight can be given to his rate of work. This criticism is, however, a minor one. In some subjects the rate of work is an important consideration but in others, particularly those in which reasoning predominates in answering the questions, the neglect of the rate of work will affect the accuracy of the examination marks only slightly, if at all.

---

[2] Comin, Robert. "Teachers' estimates of the abilities of pupils," School and Society, 3:67-70, January 8, 1916.

[3] Charters, W. W. "Constructing a language and grammar scale," Journal of Educational Research, 1:249-58, April, 1920.

Monroe, Walter S. "The description of the performances of pupils on exercises of varying difficulty," School and Society, 15:341-43, March, 1922.

5. Written examinations are usually so short that they do not offer an adequate opportunity for a student to demonstrate his ability. This criticism is frequently expressed in the statement that it is unjust to base a student's standing for a semester or a year on an examination paper written during a brief examination period. When stated in this way the criticism refers to two issues between which there is failure to distinguish. The first is in regard to the weight allowed the examination grades in determining a student's final standing. This question of the weight given the final examination is discussed in a later chapter, but it may be said here that the usual practise in high schools is to count the written examination as one-third of the student's total grade. The second refers to the inaccuracy of the grade due to the limited opportunity which is given the student to demonstrate his ability. For practical reasons it is necessary that measurement of the total achievement for the term be based upon a sample. In general, increasing the scope of the examination will tend to increase the accuracy of the measures yielded. Some evidence with reference to the reliability of examination grades based upon short samples will be presented in Chapter IV. It is possible for a teacher to make examinations more comprehensive. This can be accomplished in part by exercising more care in the preparation of the questions. The "new examination" in which pupils are required to do little or no writing affords one means for covering a wide range of subject-matter in a brief period. This method of improving examinations will be discussed in Chapter VI.

The final point to be made with reference to the inaccuracy of examination marks refers to the distinction between a "score" which describes a pupil's performance on the examination and a "grade" which interprets this score with reference to a norm. Failure to recognize this distinction is primarily responsible for too high grading by some teachers and too low by others. Even the same teacher is likely to assign "high grades" on some examinations and "low grades" on others.

In order to understand how norms (standards) are used in connection with the grading of examination papers it is necessary to distinguish between "scores," or measures, and "grades," or marks. A "score" simply describes the performance which has been recorded on the examination paper. For example, a pupil may answer 55 percent of the questions correctly. In this case 55 is his "score." If a certain number of points or credits had been given for each question his score might be 129, or 91, or 217. A "grade" interprets this description with reference to certain norms. A "grade" indirectly describes a pupil's performance on an examination, but it tells also whether the performance is to be considered as above or below passing; whether the pupil is to receive the highest mark or the lowest mark or an average mark.

It is customary to describe the quality of examination papers in terms of the percent of questions answered correctly. For example, if an examination includes ten questions and a pupil answers seven of them correctly and an eighth one partially right, he is given a score of 75 percent, which is interpreted to mean that in the judgment of the examiner he has answered the questions 75 percent correctly. School marks or "grades" are also frequently expressed in terms of percents. Sometimes they are expressed in terms of letters or other symbols, but these in turn are defined in terms of percents. For example, the grade of "A" may be defined as being between 95 percent and 100 percent. Since both "scores" and "grades" are generally expressed in terms of percents, it is only natural that the two have been confused and that "scores" have been used as "grades."

A good illustration of their difference came to the writer recently. An examination in mathematics was given to nearly 1000 freshmen in one of our large universities. This examination may properly be described as "hard," considering the training which the students had received. One student made a score of 100. The lowest score was 12. The average was approximately 55. From the standpoint of the distribution of scores this was a "good examination." If it had been easier, so that any considerable number of pupils received scores of 100 percent, it would have been unsatisfactory. If it had been so "hard" that a considerable number of students made zero scores it would also have been defective. In both cases it would have failed to differentiate between some students who were not equal in ability. However, obviously an injustice would be done if a passing mark of 70 or 75 were adopted and all pupils having scores below this mark were given a grade of failure. The passing mark for this particular examination should be in the neighborhood of 40. If the "scores" are to be represented in terms of "grades" a "score" of 40 should be translated into a "grade" of 70 or whatever passing mark has been adopted by the institution.

The recognition of this distinction between "scores" and "grades" enables us to indicate the way in which subjective norms are implied in "grades." A "grade" is not a pure measure or description of the pupil's performance. It is rather an interpretation of the measure of his performance with reference to certain norms. When no distinction is made and "scores" are used as "grades," pupils will receive high "grades" if the examination is "easy;" and low ones if it is "hard." Thus, the difficulty of the examination is one factor in establishing the norms with reference to which the "scores" are interpreted when they are used as "grades." Severe marking will tend to set high norms. Only when the examination is of average or "standard" difficulty and the marking is average in severity do "scores" and "grades" become identical in magnitude. Since the norms are established by the difficulty of the examination and the severity of the scoring, they must be subjective. In the investigations of the marking of examination papers it was shown that teachers varied widely in their judgments concerning the worth of examination papers. There is no reason to expect that they would agree more closely in estimating the difficulty of examinations. Hence, norms which depend upon teachers' estimates of the questions appropriate for examinations and upon their marking of the papers must be considered subjective. It is possible to increase greatly the objectivity of these norms and the first requirement is to recognize the distinction between "scores" and "grades." (See page 38 for a further consideration of this topic.)

Summary of inaccuracy of examination marks. From the preceding discussion examination marks are, without doubt, shown to be far from accurate measures of school achievement. However, it does not necessarily follow that the errors involved are of sufficient magnitude to justify the abolishment of written examinations. In the writer's belief the greatest benefit will come from making an intelligent inquiry into the nature of these errors and from taking steps to reduce them to the lowest magnitude.

II. Written examinations tend to encourage undesirable mental processes. Many critics have claimed that most examinations, particularly those given at the end of a course, tend to encourage "cramming." The assertion is made that many students do little or no studying until near the close of the term. Then by the process of "cramming" they are able to pass the final examination and attain a relatively high standing in the course. This criticism assumes that "cramming" is an undesirable mental process and that final examinations are responsible for its occurrence. The undesirable feature is the neglect of study throughout the term. This is not due to the fact that final examinations are given but that undue emphasis is placed upon them and that the teacher has failed to check up on the student's work day by day throughout the term.

One of the points which may be made in favor of final examinations is that they furnish an immediate incentive for review and organization of the content of the course. The writing of an examination itself may be an important part of the student's learning. This is particularly true in the case of questions which require reasoning and organization of information. "There is no impression without expression," and the writing of a three-hour examination is undoubtedly an intensive form of expression. Hence, one is justified in maintaining that written examinations tend more to encourage desirable mental processes than undesirable ones.

III. Passing the final examination an undesirable objective. The assertion has been made that when a final examination is required, the passing of it tends to become the objective for which many students work. When this occurs it is due not to the fact that the final examination is required but rather to the undue emphasis which is placed upon it by the school. If an examination consists of appropriate questions it is not undesirable to have the student keep it in mind as one of the objectives to be attained by studying the subject-matter of the course. However, as we shall show later, (see page 25) the usual practise is to count the final examination grade as one-third in determining a student's final standing. In many schools it receives less weight. When the final examination counts only one-third or less in determining a student's final standing it is difficult to say in what respect it forms an important educational objective.

IV. Examinations injurious to health of students. Some critics claim that written examinations, particularly those given at the end of a course, are injurious to the health of students, many of whom make very strenuous preparation for them. The obvious strain which accompanies the writing of answers to the questions of examinations sometimes lasting two or three hours must also be borne. It is undoubtedly true that both the preparation and the writing frequently make enormous drains on the energies of students. However, no careful investigation has been conducted of the actual effect upon their health. To one who observes the great expenditures of time and energy devoted to social and athletic activities, it is difficult to believe that examinations are in general more injurious to the health of students than many other activities in which they are permitted and even encouraged to engage. Here again it should be realized that this criticism is not fundamentally a criticism of examinations, but rather of setting very long examinations or of placing extreme emphasis upon them by making the final grade of the course depend wholly or very largely upon the examination grade.

V. Time devoted by teachers to written examinations not profitably spent. In the opinion of some critics the time given to the preparation of questions and particularly to the marking of examination papers might be more profitably employed. Information concerning the time actually devoted to the preparation and the administration of written examinations is given in Chapter III. However, it may be pointed out here that a teacher can not attain a high degree of efficiency as an instructor unless he checks up the work of his students in order to assist those who need supplementary and remedial instruction. Only by knowing the extent to which his students have achieved individually and collectively can a teacher make his instruction fit the needs of his class. Thus considerable time must be given to measuring the results of teaching. This is an indispensable portion of the teacher's task. It is only when a teacher devotes an undue proportion of his time to the preparation and administration of examinations that such work tends to be wasted. Doubtless, the time devoted to written examinations might in many cases be profitably increased. Students receiving low marks should have their answers studied in order to ascertain in what ways and why they have failed. Such information will frequently be exceedingly illuminating to the instructor, and aid him in determining his own shortcomings.

CHAPTER III

PREPARATION AND ADMINISTRATION OF EXAMINATIONS IN HIGH SCHOOLS

The data collected. The purpose of the study reported in this chapter was to secure information concerning the present practise in the preparation and administration of written examinations in high schools. A questionnaire was mailed in the fall of 1922 to 254 high-school principals in Illinois and a second one was sent to approximately 2900 high-school teachers.[1] One hundred and eighty-nine replies were received from principals and 1816 from teachers. Of the latter it was necessary to discard eighty so that the following report is based upon returns from only 1736 high-school teachers who are distributed as follows:

Commercial Subjects	192	Modern Languages	82
Drawing and Art	26	Music	21
English	342	Science	309
Home Economics	143	Shop Work	58
Latin	118	Social Science	198
Mathematics	247		

Representative character of data collected. The high schools from which answers to the questionnaire were received ranged from those established in rural communities to a large metropolitan high school. No supplementary investigation was made to ascertain the extent to which the replies were representative of conditions in Illinois but in the tabulations there was no indication that the data collected were not representative of the state. A few of the replies, particularly those of teachers, suggest that some slight misinterpretation of certain of the questions may have been made. (See page 23). Such cases, however, were relatively rare and probably did not affect the median of the results.

Extent of the requirement of final examinations in Illinois high schools. Evidence of the subjectivity of the marking of examination papers, together with other adverse criticisms of written examinations, has tended to cause many teachers and superintendents to be skeptical of their value. In a number of schools final examinations have been abolished or made optional with the teachers and they are not considered essential by many teachers. In order to ascertain the present practise in Illinois the high-school principals were asked, "Do you require your teachers to give final examinations?" Only twenty-one principals or 11 percent stated that final examinations were not required. Thus it is the practise in Illinois high schools to require that final examinations be given. This, however, does not mean that all students must take them. Of the 168 high schools in which final examinations are required 101 or 60 percent reported that it was their practise to exempt certain students. Scholarship, that is making a grade on daily work above a certain average, was mentioned by all of these schools as one of the conditions on which exemption was based. Deportment was mentioned by 52 percent and attendance by 32 percent as additional conditions.

---

[1] These questionnaires are reproduced in the appendix on pages 66 and 68.

No information was secured with reference to the explanation of the exemption from examinations of students meeting certain conditions but general observation has indicated that two reasons are frequently recognized. The first is that promise of exemption from the final examinations operates as a powerful motive to secure a high quality of daily work, regular attendance, and good deportment. The other is the belief held by many teachers that final examinations are unnecessary to determine a student's standing in a course. They contend that the average of a student's daily grades should be taken as a final grade for the course.

There is no doubt that the promise of exemption from the final examination operates as a powerful motive in the case of many students. It should, however, be recognized that such an incentive is artificial and therefore open to criticism. In so far as possible a student should be actuated by motives which sustain an intrinsic relation to the subject-matter. If it is necessary or advisable that the final examination be considered as a motive, it could be used to encourage systematic review and organization of the course. This should constitute a very important phase of studying. Students may, of course, be asked by their teachers to review frequently and to summarize and organize the work at the end of the term, but they cannot be convinced easily of the necessity of such work if it receives no more weight in determining their final grade than their performances during an equal period of time elsewhere in the course.

The second reason is a valid one in many cases. In the experience of most teachers the mark made on the final examination changes the standing of relatively few students. Experienced teachers can under favorable conditions estimate with considerable accuracy the achievements of their pupils. If the class is reasonably small and if the teacher has used methods of instruction which call for frequent oral and written performances by the students and has kept a careful record of these performances throughout the term, his estimates will generally be relatively accurate measures of the achievements of the students. There are, however, certain limitations which should be noted. Teachers may be unduly influenced in their estimates by the more recent performances of their students. Unless careful records have been kept throughout the term inferior work at the beginning tends to be overshadowed by good or excellent work during the closing weeks. In case the class is a large one the teacher may not have an adequate opportunity for becoming acquainted with all of its members.

Teachers' estimates are likely to be materially affected by personal characteristics of students; one with a pleasing personality is in many cases rated higher than one who is unattractive. If the classwork is conducted so that there is little or no written performance required, teachers' estimates will necessarily be based almost wholly on the oral responses given during the class period. Some pupils make a good showing in class when the recitation is oral but are at a decided disadvantage when asked to record their answers in writing. Frequently this difficulty is encountered when they are careless in their thinking and do not have clear ideas to express. In oral recitation they are able to make a fair showing because of personal characteristics and because of the stimulus of detailed questioning by the instructor. Furthermore, in a class discussion a bright student who has a good command of language may easily pick up ideas from other members of the class and recall ideas from his general experience sufficient to make a good showing. On the other hand there are students who express themselves more effectively in writing. They may be good thinkers but a little slow in their mental processes and not clever in discussion. Thus there are cases in which it is difficult or impossible for a teacher to estimate accurately the real achievements of students from their daily work alone. The final examination at the end of the term will in a considerable number of cases furnish additional information which is needed in arriving at the student's true standing.

The final examination in itself provides a distinct type of educational opportunity which does not occur elsewhere in the course. Altho the writers have no evidence to present upon this point they are convinced from their experience with college students and from the comments of a number who have been exempted from final examinations in high school that the practise deprives students of an important educational opportunity. Not infrequently students who have been "excused from examinations" in high school state that they experienced a distinct handicap when they entered college. If final examinations can be justified they should be required of all students. To use them only as a device for motivating the work of the term destroys much of their value.

Time devoted to written examinations. Three questions were asked relative to the time devoted to the preparation and administration of written examinations. The replies from the principals indicated that the most frequent practise is to allow ninety minutes for the writing of a final examination. This is the time allowed in 45 percent of the schools having final examinations. Fifteen percent allow eighty minutes and a slightly larger percent one hundred and twenty minutes.

The teachers were asked to state approximately the number of minutes which they use "in preparing questions for a final examination which students are allowed a total of ninety minutes to answer." The median time which varies only slightly for the different subjects is approximately fifty minutes. Individual teachers in the same subject differ widely in the amount of time which they give to this phase of their work. Two teachers, one in mathematics and one in science, stated that they spent more than six hours in the preparation of a set of final examination questions. In each subject there were a number of other teachers who stated that they devoted not more than thirty minutes to such work. It is possible that some teachers failed to interpret this question correctly but doubtless much of the variation is due to differences in the practises of teachers during the semester. Some probably make a memorandum of questions as they occur during the term and use this list as a basis for preparing the final examination. Also experience is a contributing factor. Teachers who have become very familiar with the subject should be able to formulate questions more quickly than those who are not so well versed.

The teachers were asked also to give the approximate time which they used in "marking the papers of a final examination which students are allowed a total of ninety minutes for answering." They were directed to base their estimates upon a class of twenty-five students. The median time is approximately two and one-half hours. The variations between the different subjects are not large when the differences in their character are considered. The greatest number of hours are required for English and social science and the least for drawing and art. Here also there were wide variations in the amount of time reported by the individual teachers. A total of twenty-five teachers in which all subjects except home economics, shop work, and social science were included, stated that they devoted not more than thirty minutes to marking a set of papers for twenty-five students. On the other hand, thirty-nine teachers stated that they spent 480 minutes or eight hours in the marking of a single set of papers.

It is obvious from the replies received that some teachers treat this phase of their work much more seriously than others or that they employ widely different methods. Probably some correct all errors or insert references which will enable the students to correct their own errors when the papers are returned. Others merely check the errors and still others probably do not attempt to even check each error but estimate the worth of the paper as a whole. The question concerning the amount of time which a teacher is justified in devoting to the marking of a set of examination papers may very profitably be raised. Final examination papers should be treated seriously and there should be an earnest endeavor on the part of the teacher to estimate as accurately as possible the grades which are assigned but it is doubtful if the expenditure of as much as twenty minutes per paper which was reported in some cases could be justified. The median practise seems to represent a more reasonable amount of time.

Characteristics noted in marking examination papers. The principals were asked to state whether it was the practise in their schools for teachers to subtract from a pupil's grade for (1) poor writing, (2) poor spelling, (3) poor English. Seventy-one principals or 42 percent stated that teachers were accustomed to make deductions for poor writing. In 60 percent of the schools it was the

TABLE I. PERCENT OF TEACHERS REPORTING INTENTIONALLY LOWERING A STUDENT'S GRADE FOR POOR WRITING, POOR SPELLING, AND POOR ENGLISH

Subject	Poor Writing	Poor Spelling	Poor English
Ancient Language	31	75	80
Commercial Subjects	61	74	72
Drawing and Art	65	82	67
English	60	96	98
Home Economics	37	72	73
Mathematics	31	48	48
Modern Language	31	82	74
Music	48	60	48
Science	35	53	56
Shop and Vocational	32	77	74
Social Science	37	61	65

practise to lower a student's grade for misspelled words and in 68 percent for poor English. Fifteen principals or 9 percent reported that all three characteristics were recognized only in the marking of papers in courses in English.

The teachers were asked if they intentionally lowered a student's grade because of each of the three characteristics mentioned in the above paragraph. A summary of their replies is given in Table I, which indicates considerable variation with reference to the influence of poor writing, poor spelling, and poor English upon examination grades. Since writing, spelling and English may be considered essential parts of courses in English we should naturally expect that teachers of this subject would intentionally lower a student's grade for defects in any of these characteristics. Outside of the subject of English, the majority of teachers do not lower a grade for poor writing except in commercial subjects, and drawing and art. With the exception of mathematics deduction is made by most teachers for poor spelling. The potency of poor English in determining a student's grade is slightly less than that of spelling in a number of subjects.

The handwriting, spelling, and quality of English which a student uses in writing an examination should be recognized. It does not, however, follow that a student's standing should be intentionally lowered for poor handwriting, poor spelling, and poor English in school subjects other than English. When this is done his grade becomes a measure of these abilities as well as of the abilities in the field of the subject in which the examination is given. In history, for example, a student's grade would become a composite measure of his achievement in history, the legibility of his handwriting, the quality of his spelling and the use of grammatically correct English. As a result both the teacher and the student are likely to be confused concerning the shortcomings of the examination paper. A better procedure would be to keep a record of the errors in spelling, poor writing, and poor English and when it is considered desirable a separate grade may be given covering these three characteristics. Credit for a course may be withheld until the student has brought his handwriting, spelling and English up to a satisfactory standing.

The weighting of questions. Sixty-four percent of the high school principals indicated that their teachers were accustomed to give more credit for correct answers to difficult questions than to easy ones. Approximately four-fifths of the teachers replying to the questionnaire stated that they attempted to weight examination questions on the basis of difficulty. Thus there is a very definite effort to eliminate the errors introduced in examination grades by the unequal difficulty of questions. (See page 10.)

Recognition of rate of work. Eighty-two percent of the teachers stated that they were accustomed to set examinations short enough so that practically all students could answer all the questions. Only 32 percent noted the time which each student spent in writing his examination paper, and only 8 percent said it was their custom to set examinations long enough so that practically no student would have time to answer all of the questions. Thus it is clear that relatively few teachers recognize the student's rate of work in determining his grade on an examination.

Incidentally it may be noted that when examinations are short enough so that practically all students can finish a great deal of time is wasted. Individual differences exist in all classes and it is not at all unusual to find some student finishing in one-third to one-half of the time which others devote to the examination. Aside from the waste of time which results from this practise, it is likely that the confusion caused by the leaving of those pupils who have finished tends to disturb the attention of those who are still writing. If final examinations constitute a valuable educational opportunity, there is no justification for wasting time. It is much better to set an examination long enough so that practically all students will be occupied for the entire period.

TABLE II. PERCENT OF TEACHERS GIVING AFFIRMATIVE ANSWERS TO FOUR QUESTIONS RELATING TO THE MARKING OF EXAMINATION PAPERS

Subject	Correct Answers Written	Each Ques. on one paper	One Ques. on all papers	Each paper as a whole
Ancient Language	9	73	24	20
Commercial Subjects	48	82	16	31
Drawing and Art	40	75	25	42
English	22	76	19	34
Home Economics	19	68	24	38
Mathematics	72	75	18	24
Modern Language	15	74	23	33
Music	35	80	21	42
Science	27	72	27	27
Shop and Vocational	36	75	27	37
Social Science	18	75	22	40

Method of marking examination papers. Scientific investigation has revealed that the reliability of examination grades can be materially increased by adopting a systematic method in marking papers.[2] Among the procedures recommended are the writing out of correct answers, and the grading of one question on all of the papers before taking up another question. In order to ascertain the practise of high-school teachers in marking examination papers the following four questions were included in the questionnaire sent to them.

1. Before starting to grade a set of examination papers do you write out the answers which you consider correct?
2. Do you usually grade all the answers on one paper before taking up those of another paper?
3. Do you usually grade the answers to one question on all of the papers before taking up the answers to a second question?
4. Instead of marking the answers to each question separately do you attempt to estimate the value of the paper as a whole?

The percent of teachers giving affirmative answers to these questions is given in Table II. A few apparent discrepancies in this table are due to the fact that certain teachers did not answer all the questions. Normally we should expect a teacher who answered the second question affirmatively to answer the third one negatively. This, however, did not always happen.

---

[2] Kelly, F. J. "Teachers' marks," Teachers College Contributions to Education, No. 66. New York: Teachers College, 1914, 83 p.

With the exception of mathematics, it is not the custom of teachers to write out the answers to their questions. No data are at hand to show what effect this has upon the accuracy of the examination marks. Experience with standardized tests would indicate that the failure to write out correct answers, at least in abbreviated form, would operate to make the grading of examination papers less accurate.

About three-fourths of the teachers are accustomed to mark all the questions on one paper before taking up another. This plan has the advantage of enabling the teacher to consider a pupil's performance as a whole. In the case of students who make a large number of errors the teacher will find this helpful in providing remedial instruction. It has, however, been proposed that the reliability of examination grades can be increased by marking the answers to one question on all the papers before taking up the answers to another question.

Directions to students concerning methods of work. The high-school teachers were asked the following question, "Do you prepare in written form carefully worded directions to the students regarding the procedure they are to follow in answering the questions? (These directions might include such points as, order in which questions are to be answered, length of answer, arrangement of work, etc.)" Only 35 percent of the teachers gave an affirmative answer. It is possible that in some classes there is a sufficiently definite understanding concerning the methods of work to be followed and explicit directions are unnecessary. However, it is likely that in many cases students would be able to give more truthful evidence of their ability if they were given precise directions concerning the length of answer, desired arrangement of work, etc. The order in which the questions are to be answered is a point which should be stressed. In the case of questions that are at all indefinite or general there should be specifications concerning the degree of elaborateness which is expected in the answer.

Recognition of a standard distribution in assigning grades to examination papers. The teachers were asked the following question, "In assigning grades to examination papers do you attempt to have their distribution conform to any standard form such as the normal distribution?" Only 31 percent of the teachers gave an affirmative answer to this question. This probably means that relatively few teachers have recognized the distinction between "scores" and "grades," (See pages 11-13 for an explanation of this distinction.) and for this reason are neglecting one means of making their grades more accurate measures of school achievement.

Relation of examination grades to final grades. The principals were asked if they advised their teachers as to the proportion of the final mark for the semester which should be based upon the final examination. Only eleven or 7 percent replied negatively. Of those who gave advice on this matter 95 percent made a definite ruling. The most frequently mentioned proportions allotted the final examination are 25, 30, 33⅓, and 40 percent. In 4 percent of the schools the examination counts for one-half in determining a student's final grade, in 1.3 percent for only one-tenth of the final grade. The teachers were also asked this same question. The replies varied from 10 percent to 50 percent. Except in science and shop work, the median practise is to estimate the final examination mark as one-third in determining a pupil's final grade. A considerable number of teachers indicated that they gave not more than one-fourth or one-fifth value to the final examination.

Summary. The typical practise with reference to final examinations in Illinois high schools may be summarized as follows:

1. Final examinations are required of students and exemptions are made largely on the basis of scholarship.

2. Students are allowed ninety minutes for writing a final examination. Teachers spend slightly less than one hour in preparing examination questions and from two to to three hours in grading a set of papers for twenty-five students.

3. With the exception of mathematics, the majority of teachers lower a student's grade for spelling, and with the exception of mathematics and music the majority lower it for poor English. Poor writing is not a potent factor except in English, commercial subjects, and drawing and art.

4. About three-fourths of the teachers attempt to weight the questions on the basis of difficulty.

5. The majority of the teachers do not consider rate of work in estimating the grade assigned the final examination paper.

6. The majority of teachers do not write out the answers to the questions preparatory to the marking of papers. The general practise is to mark all answers on one paper before taking up the next.

7. About one teacher in three writes out directions with reference to the procedure which the students shall use in answering the questions.

8. About one teacher in three tries to have his grades conform to a standard distribution.

9. The proportion of the final mark for the semester which is based upon the final examination grade varies from 10 to 50 percent. The median practise is 33⅓ percent. The majority of the principals make a definite ruling regarding the value placed upon the final examination.

CHAPTER IV

THE CONSTANT AND VARIABLE ERRORS IN EXAMINATION GRADES

Constant and variable errors of measurement.[1] Two types of errors are encountered in educational measurement. The presence of *variable errors* is indicated when a test is given twice to the same group of pupils. The two average scores of the group may be the same but this will not be true for individual pupils. A few pupils will make the same or approximately the same score on the two trials. Others will make higher scores on the second trial than on the first, while still others will make lower scores on the second trial than on the first. If we assume that the average of the two scores obtained represents an approximately true measure of a pupil's achievement then the differences between the first set of scores and the corresponding average scores would be the variable errors of the measures resulting from the first application of the test. Some of these differences approximate zero. Some of them are positive and about an equal number negative. Another set of variable errors would be obtained by using the scores secured from the second application of the test. In the case of a number of pupils the variable errors for the first application of the test will not be the same as those for the second application. Thus, as the name implies, variable errors change in magnitude from pupil to pupil within a group and also for the same pupil in a series of measurements of the same achievements.

A *constant error* is the same for all members of a group. Such an error occurs in teachers' marks where there is a tendency to grade too high or too low. It is found in the case of standardized educational tests when mistakes occur in the time allowed or when other departures are made from standard testing conditions. A constant error may be either positive or negative and it is generally different for different tests.

---

[1] For a more detailed discussion of the nature and magnitude of the constant and variable errors of educational measurement, see Monroe, Walter S. "The constant and variable errors of educational measurements." University of Illinois Bulletin, Vol. 21, No. 10, Bureau of Educational Research Bulletin No. 15. Urbana: University of Illinois, 1923. 30 p.

These two types of errors usually occur in combination, that is, a given measurement may and frequently does involve both a constant error and a variable error. The actual error is a combination of these two. However, in studying the accuracy of educational measurements it is helpful to distinguish between the two types and to consider each separately. The usual method used for calculating an index of the magnitude of the variable errors does not give any indication of the magnitude of the constant error. Also the method commonly used for determining the presence and probable magnitude of constant errors does not yield an index of the variable errors. Furthermore, different methods are required for decreasing the two types of errors in educational measurements.

Methods of describing the magnitude of the variable errors of measurement yielded by standardized educational tests. In describing the magnitude of variable errors in the measures yielded by standardized educational tests, the usual method is to have the test given twice to a typical group of pupils under as nearly the same conditions as possible. The coefficient of correlation between the two sets of measures is taken as the index of the magnitude of the variable errors. Usually there will be a constant error in one and sometimes in both of the sets of measures but the statistical procedure used is such that this error does not affect in any way the coefficient of correlation secured. This coefficient of correlation is commonly spoken of as the coefficient of reliability. A coefficient of 1.00 would mean that the variable errors were zero.

Available data with reference to the magnitude of errors of examination grades and standardized test scores not comparable. Investigations of the Starch-Elliott type have proven that examination grades involve errors but the method which they employed is different from that used in studying the errors in the scores yielded by standardized educational tests. Starch and Elliott confined their efforts to a study of the subjectivity of the marking of a single examination paper. Except in the use of quality scales, as in handwriting and English Composition, the scoring of standardized educational tests has been made highly objective. Hence, there has been little need for studying the subjectivity of the marking of test papers. On the other hand, there has not been, so far as the writers are aware, any reported attempt to apply to written examinations the method commonly used in studying the reliability of standardized educational tests. Hence, comparisons to show the relative reliability of the two types of measuring instruments cannot as yet be made.

For this reason in the present investigation it has seemed worth while to apply to written examinations the same method which is commonly used in studying the reliability of educational tests. Certain modifications are of course necessary. These will be noted in the following paragraphs. The investigation pertains primarily to the variable errors involved in examination grades. Incidentally some light will be thrown upon the magnitude of constant errors.

Methods employed in the present investigation of the reliability of written examinations. The essential feature of the methods employed in the present investigation is securing two independent examination grades for each pupil for the same units of work. This requires that two examinations be given to each of the groups of students from which data were secured. Two methods were used. These are described in the following directions which were sent to those cooperating in this investigation.

METHOD I

Two sets of examination questions are to be prepared by a single person, or two or more persons working together. Each of the two lists should contain the same number of questions. There should be a distinct effort to make the two lists approximately equal in difficulty and as nearly as possible similar in respect to the type of questions.

After the two lists of questions have been made both should be given by each teacher to all of her pupils under as nearly the same conditions as possible. If not given on the same day, the two examinations should be given within a period of one week. For example, if two sets of examination questions in seventh grade geography have been prepared, both sets of questions should be given by each seventh grade teacher to all of her pupils.

Each teacher is to mark both sets of examination papers for her pupils. In marking these papers the teacher should indicate the credit given for each question and write the total grade plainly upon the examination paper. When two or more teachers have given the same examinations it is not necessary that they confer in regard to the marking of the papers. If this is done a memorandum regarding the procedure should be attached to the examination papers.

This method may be followed by a single teacher who has two or more sections of a given subject. Both examinations should be given to all sections taught by this teacher. This method of studying the reliability of written examinations can be applied to any school subject. The Bureau of Educational Research is most interested in having it applied to arithmetic, history, geography, and language in the elementary school and to history, mathematics, English, and science in the high school.

METHOD II

Two sets of examination questions for the same subject are to be prepared by two teachers working independently, each teacher preparing a set. There is no requirement concerning the length or the difficulty of the two sets of examination questions except that both should cover the same amount of work. The teachers who prepare the questions should not confer concerning either the kind or the number of questions asked.

After the questions have been prepared, both sets are to be given by each teacher to all of her pupils. If not given on the same day, the two examinations should be given within a period of one week.

After the examinations have been given each teacher will grade all of the papers written upon the questions that she prepared. This will mean that she will grade a set of papers for her own pupils and also a set for the pupils of the other teacher. There should be no conferring between the teachers in regard to the method of scoring. In marking these papers the teacher should indicate the credit given for each question and write the total grade plainly upon the examination papers.

The data collected. Through the city superintendents and high-school principals a general invitation was extended to school systems in Illinois to participate in this investigation at the close of the second semester 1921-22 and also at the close of the first semester 1922-23. No instructions other than those just noted were given to those who cooperated. It should, therefore, be borne in mind that the data collected are for written examinations as they are usually given and not for special types of examinations or for unusual methods in the administration or the grading of the test papers. The reliability of examination grades could probably have been increased if certain directions had been formulated in regard to the marking of the examination papers but the purpose of the investigation was to determine the reliability of typical written examinations administered in the usual way.

Returns were secured from seventy-two groups of children but it was necessary to discard the data for six groups because instructions had not been followed. The examinations given to the sixty-six groups were all of the traditional type. The papers were marked by the teachers on the scale of 100 percent, and were then sent to the Bureau of Educational Research. The coefficients of reliability reported in this chapter were calculated under the direction of the writers.

Coefficients of reliability of examination grades. The coefficients of reliability of written examinations for the sixty-six groups of students are summarized in Table III. This table also shows the number of students in each class, the number of questions in each examination and the method followed in giving the examination. The reliability coefficients have been grouped by subjects and have been arranged in descending order of magnitude. For those entries marked with an asterisk (*) in the column headed "Method," one of the examinations was given by the principal or some other person not actually teaching a class in the subject at that time. However, this person was considered competent to be in charge of the examination. The total distribution of reliability coefficients is given in Table IV.

Two of the coefficients of reliability are negative. The highest is .95. It is interesting to note that the coefficients given for history are, on the average, higher than those obtained for arithmetic. The most reliable examinations given were in algebra. With the exception of history, arithmetic, and algebra, the number of groups is so small that comparisons can not have much significance. The median coefficient of reliability .65 may be used as a general index of the reliability of written examinations.

The coefficients of reliability of standardized educational tests. McCall[2] has stated that the "range of self-correlation for many standardized tests is about .5 to about .9." The writer's experience has indicated a somewhat greater range. In Table V the reliability of a number of standardized educational tests is given. Those for the silent reading tests by Brown, Starch and Courtis are taken from a recent bulletin[3] by the writer. The range in this table is from .19 to .92.

---

[2] McCall, W. A. How to Measure in Education. New York: The Macmillan Company, 1922, p. 396.

[3] Monroe, Walter S. "A critical study of certain silent reading tests." University of Illinois Bulletin, Vol. 19, No. 22, Bureau of Educational Research Bulletin No. 8. Urbana: University of Illinois, 1922, p.33-34.

TABLE III. COEFFICIENT OF RELIABILITY FOR WRITTEN EXAMINATIONS SET BY TEACHERS

Class Number	No. of Pupils	No. of Questions	Coefficient Correlation	Method
64 (Arithmetic)	21	10	.76	II
58 “	35	7 and 8	.74	II
43 “	24	10	.73	II
63 “	38	10	.71	II
61 “	64	10	.69	II
2 “	37	5	.67	I
62 “	38	10	.64	I
1 “	88	10	.64	I
60 “	41	5 and 10	.61	II
65 “	33	10	.60	II
4 “	72	6	.56	I
71 “	17	10	.48	*
44 “	22	7 and 10	.48	II
3 “	55	6	.47	I
66 “	73	10	.47	II
23 “	27	5	.33	I
32 “	56	5	.30	II
59 “	21	6 and 10	.29	I
45 “	43	10	.06	II
42 “	27	8 and 10	—.18	II
5 (Algebra)	51	10	.92	I
72 “	20	30 and 19	.91	*
50 “	23	5 and 7	.88	II
34 “	37	10 and 6	.82	II
57 “	36	10	.81	II
6 “	52	6	.78	—
35 “	45	15	.73	II
31 “	54	6	.61	II
7 (Language)	33	5	.68	—
9 “	54	5	.62	—
8 “	53	5	.57	—
10 (Literature)	74	5	.19	I
18 (English)	15	3	.78	I
12 “	43	5	.53	II
11 “	27	5	.53	I
54 “	37	7 and 45	.50	II
53 “	21	6 and 7	.47	II
13 “	52	5 and 8	.41	II

TABLE III. (Continued) COEFFICIENT OF RELIABILITY FOR WRITTEN EXAMINATIONS SET BY TEACHERS

Class Number	No. of Pupils	No. of Questions	Coefficient Correlation	Method
20 (History)	14	5	.95	—
15 "	28	5	.85	I
23 "	19	5	.78	I
36 "	43	5	.76	II
56 "	19	9	.75	*
67 "	53	5	.75	II
41 "	32	8 and 10	.67	II
40 "	28	5 and 7	.66	II
14 "	64	5	.63	I
48 "	24	10	.57	II
49 "	30	5 and 10	.55	II
39 (Geography)	29	10	.66	II
46 "	29	5 and 10	.66	II
47 "	47	5 and 10	.62	II
16 "	21	5	.43	I
68 "	23	10 and 8	.32	*
38 "	26	10 and 5	—.11	II
17 (Civics)	63	5	.39	I
37 (Latin)	31	8	.89	II
19 "	30	2	.82	I
51 "	50	6 and 8	.68	II
21 (Spelling)	33	5	.87	I
52 (Geometry)	42	5	.75	II
70 "	10	8	.68	*
22 "	46	4 and 5	.34	—
24 (Spanish)	18	5	.79	I
33 (German)	23	7 and 8	.83	I
69 (Commerce)	16	16 and 10	.53	*

TABLE IV. SUMMARY DISTRIBUTION OF COEFFICIENTS OF RELIABILITY FOR WRITTEN EXAMINATIONS

Size of Coefficient of Correlation	Frequency
.95	1
.90	2
.85	4
.80	4
.75	9
.70	4
.65	9
.60	8
.55	4
.50	4
.45	5
.40	2
.35	1
.30	4
.25	1
.20	0
.15	1
.10	0
.5	1
0	0
— .5	0
—.10	0
—.15	1
—.20	1
Total.........	.66
Median.......	.65

From certain unpublished studies by the writer the following information has been obtained. The Courtis Standard Research Test, Series B, Forms 1 and 2 were given to pupils as follows: Grade V, 89; Grade VI, 81; Grade VII, 52; and Grade VIII, 38. The thirty-two coefficients of reliability ranged from .409 to .904 with the median at .665. Forms 1 and 3 were given to a slightly larger group in each of the four grades. The thirty-two coefficients of correlation between the two sets of scores for this administration of Series B ranged from .528 to .963 with the median at .704. The Woody Arithmetic Scales, Series A, were given to several groups of pupils. Two scores were secured by using alternate items of each of the scales and applying Brown's formula.[4] The twelve coefficients of reliability computed in this way ranged from .91 to .46 with the average at .66. Forms 1 and

---

[4] $r_{12} = \frac{2r_h}{1+r_h}$ In this formula r_h is the correlation between two scores which this test yields. One is based upon reproduction and the other upon answers to questions.

TABLE V. RELIABILITY COEFFICIENTS OF STANDARDIZED EDUCATIONAL TESTS

Test	Coefficient
Illinois General Intelligence Scale*	.92
Courtis Standard Research Tests, Series B‡	.87
Brown Silent Reading Test—Rate	.86
Courtis Silent Reading Test, No. 2—Rate	.85
Otis Group Intelligence Scale§	.84
Monroe Standardized Silent Reading Test Revised*—Rate	.84
Courtis Silent Reading Test, No. 2—Comprehension—No. Quest	.80
Starch Silent Reading Test—Comprehension—Words	.77
Monroe General Survey Scale in Arithmetic*	.76
Monroe Standardized Silent Reading Test Revised*—Comprehension	.76
Monroe Standardized Silent Reading Test Revised*—Rate	.75
Monroe Standardized Silent Reading Test Revised*—Comprehension	.72
Starch Silent Reading Test—Comprehension—Ideas	.72
Indiana Attainment Scale, No. 1‡	.66
Starch Silent Reading Test—Rate	.62
Pressey Primer Scale†	.59
Courtis Silent Reading Test, No. 2—Comprehension—Index	.58
Pressey First Grade Vocabulary Scale‡	.37
Brown Silent Reading Test—Comprehension—Quantity	.36
Pressey Primer Scale†	.33
Brown Silent Reading Test—Comprehension—Quality	.19

*Monroe, Walter S. "The Illinois Examination." University of Illinois Bulletin, Vol. 19, No. 9, Bureau of Educational Research Bulletin No. 6. Urbana: University of Illinois, 1921, p. 47.

†Pressey, L. W. "A Group Scale of Intelligence for Use in the First Three Grades: its validity and reliability," Journal of Educational Research, 1:285-94, April, 1920.

††Unpublished data of the Bureau of Educational Research, University of Illinois.

§Colvin, S. S. "Some recent results obtained from the Otis Group Intelligence Scale," Journal of Educational Research, 3:1-12, January, 1921.

2 of Monroe's Standardized Reasoning Test in Arithmetic were given to pupils as follows: Grade V, 36; Grade VI, 92; Grade VII, 76; Grade VIII, 81. The coefficients of reliability for correct principle were as follows: .530, .630, .645, and .723; for correct answer they were .518, .528, .576, and .707. Using Brown's formula the coefficients of reliability for Gray's Silent Reading Tests were computed for thirty grade groups. These coefficients ranged from .55 to .85 with the median at .67. The number of pupils per group was less than 100 in only five cases. For several grade groups reliability coefficients were secured for Monroe's Standardized Silent Reading Tests which ranged from .222 to .907 with an average of .669.

Haggerty has computed the reliability for both Sigma 1 and Sigma 3 of his Reading Examination by having the same test repeated. In the case of Sigma 1 the interval between the two applications of the test was six weeks. For 200 children in Grades I to III inclusive the coefficient of reliability .84 was obtained. In the case of Sigma 3 the interval between the two applications was only two days. For 126 pupils from Grades V to VIII, inclusive, the coefficient of reliability was found to be .885. For the sentence test alone the reliability coefficient was .769 and for the paragraph test, .806. For Thorndike's Scale Alpha for the Understanding of Sentences, McCall has reported a coefficient of reliability of .37. This was obtained by using a test similar to Alpha but not considered a duplicate form. Gates[5] reported reliability coefficients for Thorndike-McCall Reading Scale which ranged from .25 to .72. All of these were for pupils belonging to a single grade. For the Burgess Picture Supplement Scale the author has given coefficients of reliability ranging from .62 to .88 for grade groups from the second to sixth grades inclusive. In each case the number of pupils was relatively small. Gates gave coefficients of .62, .59 and .66 for three grade groups.

For the Otis Self-Administering Test of Mental Ability the author has reported an average reliability coefficient of .921 for the higher examination and of .948 for the intermediate examination. Presumably these coefficients are based on the scores secured from pupils for a sequence of several grades. For the separate tests of the Stanford Achievement Test the authors reported coefficients of reliability based upon separate grade groups which ranged from .75 to .96. When the composite score of all the tests was used the reliability coefficient was .98.

The relative reliability of written examinations and standardized educational tests. The data which have just been submitted indicate that the difference between the reliability of the two types of instruments is not as great as is commonly believed. The median of the reliability coefficient for written examinations given in Table IV is .65. There are many reliability coefficients for standardized tests in Table V which are less than this. Furthermore, the additional citations of coefficients of correlation in the above paragraphs indicate that for a number of standardized educational tests which have been very widely used the median of the reliability coefficients for grade groups is in the neighborhood of .65. Thus the conclusion seems justified that altho some of our more elaborate standardized tests, such as the Stanford Achievement Test, the Illinois General Intelligence Scale, and the Otis Self-Administering Test of Mental Ability, may be expected to yield measures whose reliability is greatly in excess of that of typical written examinations, many widely used standardized educational tests yield measures which possess about the same degree of reliability as the grades obtained from written examinations prepared by teachers and other school officials. It should be noted that reliability refers only to the variable errors of measurement. The constant errors as we shall show, (p. 40) are likely to be very much larger in examination grades than in the scores yielded by standardized educational tests. It should also be noted that the time required to give many of the standardized tests is much less than that devoted to a typical written examination.

---

[5] Gates, Arthur I. "An experimental statistical study of reading tests," Journal of Educational Psychology, 12:379, October, 1921.

The absolute reliability of examination grades. The statement that the reliability of a typical examination is equivalent to that of many standardized tests and only slightly less than that of a number of others still leaves a doubt with reference to the absolute reliability. For practical purposes the reliability coefficient of .65 needs to be interpreted in terms of the variable errors of measurement to be expected. The correlation tables for eight groups having a reliability coefficient of approximately .65 were taken and the scores translated into a five point system of school grades. It is assumed that these classes were typical and the highest scores were translated into a mark of "A," the lowest into a mark of "E." This was done in an arbitrary way but the results indicate roughly one meaning which may be attached to a reliability coefficient of .65. It was found that in 40 percent of the cases the students received the same grade in the two examinations. In an additional 42 percent the grade which they received on the first examination was only one point higher or lower than that received on the second. For example, if a student in this group made a "D" on one examination, he made an "E" or "C" on the other. The two grades received by the remaining 18 percent differed by two points or more.

Conditions tending to produce variable errors of measurement in examination grades. Several sets of examination papers were examined in order to ascertain the conditions which tended to produce the lowest coefficients of reliability and hence the largest variable errors of measurement. The most potent cause appeared to be that the two teachers recognized widely different educational objectives in making out the two sets of examination questions. This seemed to be the case in Group 42, arithmetic, for which the coefficient of correlation was −.18. In Group 22, geometry, there was a difference in the general plan of the examinations; one teacher permitted the students to choose one of two questions in part of the examination while the other required that all questions be answered. This difference in the plan of the examination appeared to increase the variable errors of measurement. There was also a difference in the educational objectives recognized in that one teacher placed much more emphasis upon the practical application of geometry than the other.

Another cause which operated to lower the degree of correlation and hence to increase the magnitude of the variable error was the adherence to different standards of excellence by the teachers who graded the papers. For example, in Group 45, arithmetic, one teacher considered only the final answer to the exercise; if that was right the student received full credit—if wrong, no credit was given. The other teacher gave credit for correct principle. The coefficient of reliability for this group was .06.

It was noticed that in general pupils made higher grades on the tests set by their own teacher than on those set by another person. This appeared to be true even when distinct differences could not be identified either in the educational objectives or in the methods of grading of the two teachers. In Group 32 for which a reliability coefficient of .30 was obtained when the grades made on the first examination were correlated with those made on the second examination, a second coefficient was calculated by comparing the student's grade made on the examination prepared by his own teacher with that set by another teacher. This procedure gave a coefficient of correlation of .40. When the two classes were taken separately coefficients of .57 and .44 were obtained. These data tend to supplement the evidence already cited that differences in the content of the examination and in the plan of marking are potent factors in producing the variable errors of measurement.

The magnitude of constant errors in examination grades. It is probable that most of the teachers marking the examination papers did not recognize the distinction between "scores" and "grades"[6] and that the marks placed upon the papers were con-

---

[6] See page 11 for a statement of this distinction.

TABLE VI. DISTRIBUTION OF DIFFERENCES BETWEEN AVERAGES OF EXAMINATION GRADES

Difference	Frequency
50	1
32	1
29	1
27	1
22	1
21	1
20	2
18	3
16	1
15	1
14	2
13	2
12	1
10	2
9	1
8	4
7	3
6	6
5	4
4	6
3	5
2	6
1	8
0	3
Total.........	66
Median.......	6.2

sidered as "grades." In several instances the "grades" made on one examination were on the average much higher than those made on the other. If "scores" were used as "grades," any differences between the averages of the two sets of measures indicate the presence of constant errors. In order to secure an index of their magnitude the differences were calculated for the sixty-six groups to which two examinations were given. These are assembled in Table VI. For three of these groups the difference between the averages of the two sets of "grades" was zero; for eight other groups it was one. At the other extreme we find a difference of 50 in the case of one group. The median difference is 6.2.

It should be noted that the differences between the averages of two sets of examination grades are not constant errors. They are merely indicative of the presence of constant errors. If one examination was easy and the other hard the difference would be the sum of a positive error and a negative error. If both examinations were hard the difference would be smaller than the constant error in either average. The large differences shown in Table VI are probably caused by the combination of an easy examination with a difficult one. This was very obviously true in the case of the one difference of 50. Furthermore, in interpreting Table VI it should be remembered that possibly some of the teachers recognized the distinction between "scores" and "grades," and the marks would have been appropriately adjusted before being used as "grades."

So far as it was possible to ascertain from an analysis of the examination papers the large differences are due to two causes—differences in the difficulty of the two sets of examination questions and in the severity of the grading. For example one of the examinations which produced a difference of 40 consisted of seven questions of which the pupils were permitted to answer any five. These questions were relatively easy. In the other examination, there were ten questions and the pupils were required to answer all of them. Very few were able to complete this second examination in the time allowed and the teacher appears to have counted the unfinished exercises as failures. Nine out of twenty-two children in the second group made zero on the examination. In this way a very large constant error was introduced but the coefficient of reliability for this group was .48.

Relative magnitude of constant errors in examination grades and in standardized test scores. In another place[7] the writer has discussed the magnitude of the constant errors in educational tests. In cases where there has been coaching for tests, intentional or not, or disregard for standard directions, large constant errors have been introduced. In one extreme instance a constant error of over three and a half years occurred in the mental age scores of a group of children. In general, however, because of the standard directions for administering the tests and scoring the papers, of the objectivity of the marking, and of the norms for interpreting test scores, the constant errors in standardized tests are very much smaller, and are likely always to be smaller than those found in examinations given by teachers. However, some reduction in the magnitude of the constant errors in examination scores will result when the use of either very easy or very difficult sets of questions is avoided and when a conservative plan of marking is followed.

---

[7] Monroe, Walter S. "The constant and variable errors of educational measurements." University of Illinois Bulletin, Vol. 21, No. 10, Bureau of Educational Research Bulletin No. 15. Urbana: University of Illinois, 1923, p.19-20.

Explanation of the apparent contradiction between the results of this investigation and previous studies of examination grades. The results of this investigation have caused the writers to revise their estimate of the accuracy of examination grades. The findings indicate that the errors are much less than they appeared to be from evidence resulting from investigations of the Starch-Elliott type. One naturally asks the question, "Why this apparent contradiction?" Starch and Elliott obtained similar results for three different examination papers and numerous other investigators have corroborated their findings. The mass of evidence accumulated is so extensive and uniform in character that one would naturally be inclined to accept the conclusions indicated in preference to the apparent contradictory results of the present investigation. However, a careful analysis of the procedures reveals that the results are not necessarily contradictory. The method followed by Starch and Elliott combines both constant errors and variable errors. The "grades" assigned to the examination paper in geometry were influenced both by the subjectivity of the marking and by the tendency of some teachers to grade high and of others to grade low. The present investigation has separated the variable errors from the constant. It has also shown that the examination scores have in some cases involved relatively large constant errors. The extreme differences between the grades assigned to the same paper reported by Starch and Elliott (see page 9) are easily explained when it is understood that they represent the combination of variable errors and constant errors. Especially is this true when we realize that the constant errors would likely be larger for teachers of different schools as in their investigation than for teachers in the same school as in the present investigation.

Conclusion with reference to relative accuracy of examination grades and scores yielded by standardized tests. As already indicated the writers believe that the data presented in this chapter show that examination grades are more accurate measures of achievement than many persons have considered them to be. Standardized tests yield scores involving errors, both constant and variable, but in the case of our best standardized tests these errors are distinctly less than the corresponding errors in examination grades. Furthermore, measurement by means of standardized tests usually requires much less time than is commonly devoted to written examinations. This conclusion refers to written examinations of the traditional type and administered under typical conditions. It is likely that written examinations and their administration may be improved so that the difference in the accuracy of examination grades and test scores will become much less than at present.[8]

---

[8] The conditions of standardized tests would have been more closely approximated if both sets of examination questions had been prepared by the same person and marked by different persons. If this had been done it is reasonable to expect that the coefficients of reliability would have been somewhat higher and the differences in the averages of the two sets of scores smaller.

CHAPTER V

THE CONTENT OF WRITTEN EXAMINATIONS

The data collected. In response to an invitation sent to superintendents and high-school principals in Illinois sets of examinations were received from fifty-six schools for the first semester and from fifty schools for the second semester of the school year 1921-22. A range of approximately sixty subjects was represented. It seemed desirable to restrict this analysis of sets of questions to the thirteen subjects listed in Table VII. The number of sets of questions and also the total number of questions are given in this table.

Classification of questions according to type. After considerable experimentation a list of fifty types of questions as given below was formulated.

- Aims
- Analysis
- Cause (give)
- Classification
- Comparison
- Completion
- Conjugation
- Construction (a figure, study or statement)
- Construction (give the)
- Contrast (general)
- Contrast (specific basis)
- Correction
- Criticism
- Decision (choice or preference)
- Declension
- Definition
- Description (characterization)
- Diagram (illustrate by)
- Discussion
- Effect (give the)
- Evaluation
- Example (illustrate by)
- Expansion
- Explanation (tell why or how)
- Facts (definite number)
- Facts (indefinite number)
- Factoring
- How many (tell)
- Law (give the)
- Mathematical operations of addition, subtraction, multiplication, and division
- Method
- Outline
- Parsing
- Proof
- Punctuation (capitalize and correct sentences)
- Recall
- Reduction to lowest terms
- Relationships (give the)
- Rule
- Scanning
- Simplification
- Source
- Substitution (values for letters)
- Summary
- Solving for unknown quantity
- Syllabus
- Translation (foreign language to English)
- Translation (English to foreign language)
- Use (give the)
- Where (tell)

TABLE VII. NUMBER OF QUESTIONS AND SETS OF QUESTIONS EXAMINED

Subject	Sets	Questions
English I	80	721
English II	83	694
English III	79	721
Algebra I	80	731
Plane Geometry	80	636
Latin I	81	683
Latin II	76	539
Physics	76	795
General Science	62	789
Civics	59	560
American History	62	550
Domestic Science	42	392
Domestic Art	41	368
Total	901	7621

All questions for the thirteen subjects mentioned in Table VII were classified under some one of these types. This classification was made by Mr. Souders with the assistance of a single clerk working under his immediate direction. Altho any classification of this kind is necessarily subjective, a relatively high degree of uniformity has, we believe, been secured.

Summary of classification. Twenty-six of the fifty types of questions were represented in six or more of the thirteen subjects. The relative frequency of each is given in Table VIII. This classification of examination questions shows a high frequency of certain types and very little or no use of a number of other types. If we omit Latin, Algebra, and Plane Geometry in which the nature of the subject-matter restricts the kind of question asked, we find that 32 percent of all the questions require "explanation." The next most frequent type used, 21 percent, calls for a "definite number of facts."

Frequently all questions are considered as belonging to one of two groups, "thought questions" or "memory questions." Such a definite classification is not, however, always possible. The character of the mental process involved in answering depends upon the person replying as well as upon the form of the question asked. Those questions calling for definite facts are almost certain to be based upon memory; on the other hand, those requiring classification, evaluation, contrast, etc. are likely to demand

TABLE VIII. PERCENT OF RELATIVE FREQUENCY OF TYPES OF QUESTIONS IN THIRTEEN SCHOOL SUBJECTS

Type	English I	English II	English III	American History	Civics	Latin I	Latin II	Algebra I	Plane Geometry	General Science	Physics	Domestic Science	Domestic Art
1. Analysis	1	—	—	—	—	—	—					—	
2. Cause or effect	1	—	—	4	—		1			2	1	1	1
3. Classification	1	1	1	—	1	—	1		—	1	—	2	1
4. Completion	2	2	1	—		—	—	—	1	—	—	—	2
5. Construction	14	13	13	1	1	2	5	1	12	2	2	3	10
6. Contrast (Gen.)	—	—	—		—	—					—	—	—
7. Contrast (Spec.)	2	4	5	3	4	2	1	—		3	4	3	3
8. Correction	3	3	3				—	—				—	
9. Decision	2	1	—	—	1				—	—		1	—
10. Definition	5	6	4	1	6	—	1	2	5	4	11	2	2
11. Description	6	3	6	2	6		2			6	5	5	6
12. Diagram	—	—	—				—	—	1	5	6	1	3
13. Discussion	1	3	7	16	7		1			4	5	7	3
14. Effect	—	—	1	3	1		1			2	—	2	1
15. Evaluation	—	—	—	—	1			—				—	—
16. Example	5	4	3	1	2	3	2	2	1	5	4	1	3
17. Explanation	21	22	15	36	38	8	13	1	6	42	37	34	28
18. Facts (Def. no.)	18	20	20	21	20	9	7	1	4	17	10	25	29
19. Facts (Indef. no.)	1	1	2	3	2	—	—	—	—	2	1	3	2
20. How many				—	—	—		—	—		—		
21. Outline	2	2	4	5	4	—	1		—	1	—	2	3
22. Proof	—	—	—					2	56	—	1		—
23. Recall	8	7	6	3	2	1	3			1	—	2	—
24. Rule	1	1	—		3	2	3	4	2	—	2	1	1
25. Use	1	2	1	—	—	3	1				1	2	2
26. Where	1	1	—	—	—	—			1		—	—	1
All other types	5	3	5	—	1	68	60	86	11	1	12	2	
Total Questions	922	906	802	601	761	760	578	646	699	667	887	500	580
Percent of Memory	21	19	16	4	11	73	65	69	74	9	24	9	6
Percent of Thought	79	81	84	96	89	27	35	31	26	91	76	91	94

The dash (—) means a frequency of less than ½ of 1 percent.

thought on the part of most students. If, however, such classifications or evaluations have been made in a previous class exercise some students may easily remember the answers and, in such a case, a thought question for one student becomes a memory question for another. For the purpose of this study Types 10, 20, 22, 25, and 26 have been designated as probable memory questions, the remaining types as probable thought questions. The percent of each group is given in the last two lines of Table VIII. These percents can be considered as only a rough indication of the relative frequency of these two very general divisions.

In her investigation of "the question as a measure of efficiency in instruction," Dr. Stevens[1] attempted to determine the relative number of thought questions and memory questions asked by teachers in a single class period. The percents of memory questions for history, English and science were 83, 55, and 67 respectively. This relative frequency is much larger than indicated in Table VIII. The difference may be due to the fact that in the present investigation only written examination questions were considered, but it is altogether likely that it is indicative of a real change in the type of questions which teachers commonly ask of their students.

Relation of questions to educational objectives. The questions which teachers ask during class periods constitute a concrete expression of the educational objectives which they are day by day setting for their students. The questions of the final examinations should, therefore, be representative of the types of educational objectives set in the different school subjects.[2] The emphasis upon memory and some of the simpler types suggests a need for a modification in emphasis in most of the school subjects.

Quality of examination questions. Altho the writers have no objective evidence to present in regard to the quality of examination questions, those submitted for this study were in general considered good. Catch questions or those stated so that they would not be understood easily by students were very rare. Many questions were stated so that the grading of the answers was objective and would indicate that their form had been influenced by the exercises of standardized tests. This was especially true of those examinations having questions of the true-false or recognition type. In the judgment of the writers the criticism that teachers are inclined to ask misleading or catch questions is not a valid one in the case of the examinations studied in this investigation.

---

[1] Stevens, Romiett. "The question as a measure of efficiency in instruction." Teachers College Contributions to Education, No. 48. New York: Teachers College, Columbia University, 1912.

[2] See page 55 for a further discussion of objectives in school subjects.

CHAPTER VI

THE IMPROVEMENT OF WRITTEN EXAMINATIONS

Altho we now have a number of standardized tests which are superior to written examinations, and we have reason to believe that they will be used even more extensively than at present, there is need to give attention to the improvement of written examinations. It does not appear likely that standardized tests will ever replace written examinations. The latter type of measuring instrument will probably continue to be the most frequently used means of measuring the achievements of school children.

Written examinations may be improved by correcting the faults which have been noted in the preceding chapter. In this chapter we shall consider four important improvements: (1) Reduction of constant errors; (2) Reduction of variable errors; (3) Securing a greater agreement of the content of examinations with recognized educational objectives; (4) Simplification of the administration of written examinations.

There is some overlapping between these improvements. For example, the magnitude of errors in measurement, particularly variable errors of measurement, will be reduced by securing a greater agreement between the content of the examination and recognized educational objectives. The devices for simplifying the administration of examinations also tend to make the results more accurate.

Causes of constant errors in examination grades. The fundamental cause of constant errors in examination grades, i.e., "high grades" or "low grades," is the failure to recognize the distinction between "scores" and "grades." (See page 11.) A pupil's grade tells his standing with reference to a norm, i.e., the passing mark. When no distinction is made between "scores" and "grades" this norm is subjective. Altho the passing mark may be defined numerically as 70 percent or 85 percent it is fixed in the case of a particular examination by the difficulty of the questions and by the severity of the marking of the papers. Pupils will receive "high grades" when the examination is easy or the plan of marking is generous. They will receive "low grades" when the examination is hard and a severe plan of marking is followed. If the teacher makes no distinction between "scores" and "grades" he sets the norm for a particular examination when he makes out the questions and decides upon the plan of marking. He implicitly expresses the opinion that the pupil whose achievements are barely "passing" will make a grade of 70, or the passing mark adopted. He also implies that the pupil whose achievements are exceptionally high will make a high grade, i.e., a grade of 95 or between 95 and 100. Such expressions are merely subjective.

Since the failure to recognize the distinction between "scores" and "grades" is the cause of constant errors the plan for improvement is obvious. The papers should be marked in terms of "scores." These may be on the scale of 100 but this is not essential. In fact it will probably assist a teacher in keeping the distinction in mind if the scores are not on the scale of 100. After the papers have been marked the "scores" should be translated into "grades" by comparison with a norm in which the subjective elements are reduced to a minimum.

A standard average grade used as a norm. The simplest objective[1] norm is a standard average grade. This may be set arbitrarily but a more rational procedure would be to take the average of the grades given in a school on a particular subject during a period of several years.

The standard average grade defines the grade into which the average score of a typical class should be translated. For example, if the standard average grade is 85 and the average score in a particular class is 57 the grade corresponding to this score would be 85. In case the class is made up of poor students the average grade of the class should be below the standard average grade. If the class is unusually bright their average grade should be higher than the standard average. The translation of the average score into the appropriate corresponding grade furnishes a basis for the translation of the other scores of the group.

The procedure just outlined is necessarily crude. It is partially subjective because the determination of the general status of the class is left to the teacher. However, the teacher may use previous school records or the measures obtained from a standardized test to assist him in arriving at a partially objective estimate of the general status of the class. The use of a standard distribution instead of merely a standard average grade represents a more systematic procedure.

---

[1] The adjective "objective" is not intended to indicate perfect objectivity or even as high a degree of objectivity as we have in the case of many standardized tests. As used here it means that the norm is distinctly less subjective than the norm commonly implied in the usual examination.

1. Decreasing the magnitude of constant errors by means of a standard distribution of grades. For several years a number of educators have been urging that teachers make the distributions of their grades conform to a standard shape, i.e., that a specified percent of the members of a typical class be given a grade of A, another specified percent a grade of E, and so on for each of the marks adopted by the school.[2] A number of distributions have been recommended. For a five point system of grades several authors have recommended the following distribution, 7, 24, 38, 24, 7. Other distributions which have been advocated are 7, 18, 50, 18, 7 and 5, 15, 60, 15, 5.

The essential feature of the plan is a specification of the percent of the students of a given group who are to receive each mark rather than the particular form of distribution used. There is much evidence which indicates that the distribution of achievements of an unselected group of students approximates the normal probability curve.[3] If we assume that true measures of the achievements of an unselected group of 100 or more are distributed normally this adjustment does not fix the percent who are to receive each grade. The normal probability curve may be divided in many ways, for example, it is possible to divide the curve so that there would be 50 percent of A's, 20 percent of B's, 10 percent of C's, 10 percent of D's, and 10 percent of E's. In such a distribution a grade of A would be given to all students above the average of the class. An appropriate meaning could be stated also for each of the grades. A distribution which is symmetrical has certain advantages and one of those mentioned in the preceding paragraph is to be preferred. The particular standard distribution to be used is a matter of policy which each school should determine. Some argue that different standard distributions be adopted for the different years of the high school, some advocate different standard distributions for different school subjects. It should be noted, however, that there are certain advantages in uniformity. It would be desirable for all high schools, particularly those in a given state, to agree upon a common standard distribution and to use this for all subjects. Grades assigned in different schools can have a common meaning only when they conform to the same standard distribution.

---

[2] If grades are expressed in percents the corresponding intervals such as 95 to 100, 90 to 94, etc. would be used instead of A, B, C, etc.

[3] The normal probability curve is bell shaped and is symmetrical with the average or median as a center.

The proposal that teachers make the distributions of their grades conform to a standard shape has met with much criticism. As in any controversy there have been extremists on both sides and many of those participating have given evidence that they failed to understand clearly the nature of the proposal of its essential features. Among the advocates of the use of a standard distribution are those who have insisted that the normal probability curve explicitly defines the students who must receive A's, who must receive B's, etc. Cases have been reported of instructors who frankly admitted that a certain student deserved to receive an A but that they had used up all the A's which the distribution allowed, and that, therefore, the student must be satisfied with a B. One hears also of instructors who announce at the beginning of a course that a certain number of the class must fail. It is rumored that in some of these instances the students enrolled have hired certain other students who were indifferent to their scholastic standing to enter the course in order to provide the requisite number of failures. The opponents of the use of a standard distribution have contended that there was no a priori reason why any student should fail and that always the quality of the student's work should determine his scholastic standing. Furthermore, they have pointed out that in any group of students brought together for instructional purposes it is extremely unlikely that the distribution of achievement would approximate at all closely their predetermined standard distribution. The mechanical and unintelligent application of a standard distribution by some instructors has given the opponents of the plan concrete examples of what they imagined to be the normal result of its use.

A standard distribution is merely a device which teachers may use in order to reduce to a minimum the constant errors in their grades, but to be helpful it must be used intelligently. It must be remembered that a standard distribution is a means and not an end. Whenever common sense indicates that the distribution of grades for a particular class should depart from the normal distribution no instructor should hesitate to award the grades which he believes the students deserve. It is intended that the standard distribution will be closely approximated only for a large unselected group of students. A particular class very frequently is made up of a selected group of students. Furthermore, classes of the usual size, 20 to 35, are so small that frequently there will be significant departures from this standard distribution.

Translating scores into school marks by means of a standard distribution. A standard distribution is useful in translating examination "scores" into "grades." The examination papers should be marked in terms of a score. This score may or may not be on the scale of 100 points. In order to avoid confusion between "scores" and "grades" it is wise to use a scale of points shown so that the maximum score will not be 100. If the papers have been marked in this way the scores may be arranged in columns as indicated in the left hand margin. The first step in translating these scores into grades is to determine whether or not the class is typical. If an experienced teacher has had a class for several weeks he will usually be able to estimate its general status with a fair degree of accuracy. At the beginning of a school year or in the case of an inexperienced teacher some outside information is needed. The previous school record of the students may be studied but in many cases it will be more convenient to administer a general intelligence test. The average mental age and the distribution of the I.Q.'s of the class will be a very reliable index of the composition of the group. If the median I.Q. is distinctly below 100 the teacher may know that he has poor pupil material. If it is much above 100 he knows that the class consists of pupils better than the average. If there is an unusually high number of low I.Q.'s he may expect a relatively high number of low grades.

74	A
73	
70	
69	B
68	
64	
63	
60	
58	C
57	
56	
55	
54	
51	
50	
47	D
46	
42	
41	
40	
38	E
35	
32	

With the general status of the class in mind the scores may be grouped in conformity with the system of marks used. In the illustration in the left hand margin it has been assumed that the class is approximately typical. The percent of A's and also the percent of failures are somewhat larger than the percent specified in most standard distributions. If the scores are arranged in the form shown below the general shape of the distribution will be more obvious. However, in the majority of cases it will be sufficient to use the arrangement given in the margin.

		58		
		57		
	47	56	69	
	46	55	68	
38	42	54	64	74
35	41	51	63	73
32	40	50	60	70
E	D	C	B	A

It is seldom that one will have exactly a symmetrical distribution of grades for a class of this size. Some departures from the standard distribution must be expected. In case the class is not typical one should expect marked departure from the standard distribution. For example, the distribution for a given class might be as shown below. In this there are no grades below passing but there are a number of poor students just above the passing mark. Also the percent of A's and B's is unusually large. Such a distribution is not normal but might well represent the distribution of grades for a particular class even when the normal distribution had been adopted as the standard. If the teacher is able to show that the general status of the class justifies such a departure he deserves commendation rather than criticism for his distribution.

			64	
	42		63	
	41	54	60	74
	40	51	58	73
	38	50	57	70
	35	47	56	69
	32	46	55	68
E	D	C	B	A

An accumulative distribution used as a check upon constant errors. A standard distribution is also useful as a check upon the grades given by a teacher over a period of several terms. When the grades for the entire period are assembled in such a distribution any general tendency on the part of the teacher to give too high or too low grades will be revealed. Each teacher should keep an accumulative distribution of the grades in each subject he teaches. For example, a teacher in mathematics should keep an accumulative distribution of the grades given in classes of first-year algebra. When the total number of grades becomes large a comparison of this distribution with the standard distribution will reveal any tendency on the part of the teacher to grade too high or too low in this subject. A teacher should then take steps to correct any marked departures from the practise defined by the standard distribution. In large schools where there are several sections of the same subject it will be helpful to secure a distribution of grades each time they are issued. Any marked departures from the standard distribution will then be called to the attention of the teachers. However, one should avoid giving the impression that there must be uniformity with the standard distribution. Departures from this standard distribution are justified when the group of pupils can be shown to be selected. Thus a departure from the standard distribution is a cause for an investigation on the part of the teachers concerned. If evidence can be produced which justifies the departure no change in the system of grading should be used. On the other hand when investigation reveals no reasons why there should be departures from the standard distribution, the teachers should be urged to modify their system of grading so that a greater uniformity will be secured.

2. Decreasing the variable errors in examination scores. The reduction of the magnitude of variable errors of measurement in examination scores is to be secured mainly through the adoption of rules which will bring about greater uniformity in preparing and administering examinations. These rules should include specifications in regard to the effect of poor writing, poor spelling, and poor English upon a student's grade, and should be in agreement in regard to giving credit for correct principle and partial credit for exercises partly right or partly completed. The rules may properly include also specifications relating to the number and types of questions to be asked and the form in which they are to be presented to the students. For guidance in marking papers a teacher should write out, at least in abbreviated form, the correct answers to the questions. The accuracy of examination scores will be increased also by making the examinations more uniform with respect to content.[4]

---

[4] For recommended rules covering these and other points see Chapter VII.

It has been proposed that the use of types of questions which call for answers that may be objectively classified as either "right" or "wrong," would facilitate uniformity in marking the papers. This means of reducing the variable errors of measurement will be considered under the head of "simplifying the administration of written examinations."

3. Securing agreement of the content of examinations with recognized educational objectives. The intrinsic function of an examination is to measure certain achievements. In general the achievements for which we desire to secure measurements are those included in the recognized educational objectives. Hence, the questions should be in agreement with the objectives. Therefore, it is impossible to cover all details in a given subject-matter field. The questions should relate to the most significant facts, principles, etc. of the course. Catch questions and those calling for unimportant details have no place in an examination. For example, an examination in spelling should not include unusual or obsolete words, an examination in history should not call for obscure dates or other trivial facts.

In securing agreement the teacher should make use of such terms of minimum essentials as are available. For example, in spelling a teacher may very properly select the test words from Ayres' list of the one thousand most frequently used words or from some other carefully prepared minimum essential list. In geography a teacher will find the Hahn-Lackey Geography Scale a helpful source of questionings. In other subjects the teacher will not have access to terms of minimum essentials as complete as in these two subjects, but he should become familiar with curriculum studies and other investigations[5] relating to educational objectives.

---

[5] The following list is suggestive of studies relating to educational objectives:

Yearbooks of the National Society for the Study of Education. Bloomington, Illinois: Public School Publishing Company.

Part I of 14th—reading, writing, spelling, language and grammar, arithmetic, history, literature, geography.

Part I of 16th—reading, writing, spelling, arithmetic, history, physical education.

Part I of 17th—arithmetic, geography, reading, English, civics, history.

Part II of 17th—history, civics, economics, sociology, geography.

Part I of 19th—on new materials of instruction, reading, history, geography, mathematics, nature study, civics.

Part I of 20th—on materials of instruction—all subjects in elementary schools.

Part II of 22nd—the social studies in the elementary and secondary school.

"Arithmetic, course of study for the elementary schools, including the kindergarten and the first six grades." Course of Study Monographs, Elementary Schools, No. 1 Berkeley, California: Public Schools, 1921. 86p. *(Concluded on p. 56.)*

In some subjects there are valuable committee reports which give the consensus of opinion concerning the relative importance of the numerous distributions.

The teacher must assume most of the responsibility for securing the agreement between the content of the examination and educational objectives. In many of the high-school subjects he can obtain little assistance from such sources as just indicated. However, if this purpose is kept in mind and if he is really familiar with the subject which he is teaching, gross inconsistencies with recognized educational objectives will be avoided.

4. Simplifying the administration of written examinations. The administration of written examinations, particularly the marking of the papers, can be greatly simplified by the use of certain types of exercises. For example, in the true-false type of exercise the pupil merely indicates whether the statement is true or false. Instead of asking the question, "Why did the Puritans come to America in the seventeenth century?" we may ask whether the following statement is true or false. "The Puritans came to America in the seventeenth century seeking wealth." The pupil may give his answer to this exercise by writing a plus sign after the statement if he considers it true and a minus sign if he considers it false. In case the statement is dictated to him he may write after the number of the exercise the word "true" or "false" or the appropriate sign. The answering of such exercises requires very little of the pupil's time and the scoring is exceedingly simple. Questions which can be answered merely by "yes" or "no" also simplify the administration of examinations. Similar results can be secured with recognition exercises such as have been used in a number of standardized silent reading tests. The following is an exercise of this type.

Ayres, L. P. "A measuring scale for ability in spelling." N. Y.: Division of Education, Russell Sage Foundation, 1915. 58p.

Ayres, L. P. "Measuring scale for handwriting." N. Y.: Division of Education, Russell Sage Foundation, 1920. (Folder, chart.)

Bagley, W. C. and Rugg, H. O. "The content of American history as taught in the seventh and eighth grades." University of Illinois Bulletin, Vol. 13, No. 51. Urbana: University of Illinois.

Charters, W. W. Curriculum Construction. N. Y.: Macmillan Co., 1923. 352p.

Charters, W. W. and Miller, Edith. "A course in grammar." University of Missouri Bulletin, Vol. I, Education Series 9. Columbus: University of Missouri, 1915.

Hahn, H. H. Hahn-Lackey Geography Scale. Wayne, Nebraska: H. H. Hahn, State Normal School.

Hahn, H. H. Scale for Measuring Ability of Children in History. Wayne, Nebraska: H. H. Hahn, State Normal School.

Moore, E. C. Minimum Course of Study. N. Y.: Macmillan, 1923. 402p.

> "The first president of the United States was: Christopher Columbus, Benjamin Franklin, George Washington, Thomas Jefferson."

In answering this exercise the pupil is asked to underline or mark in some other way the name required to make a true sentence. Completion exercises in which pupils are asked to supply words which have been omitted furnish still another means of simplification.

Directions for constructing a true-false examination.[6] 1. In constructing true-false exercises, a list of statements covering in some detail the portion of the subject on which the pupils are to be examined should be prepared. Some of the statements can then easily be changed so that they are false. The untruth of a statement should not be too obvious or it will be worthless for testing. Also statements should be selected which require an acquaintance with the subject in order to determine their truth or falsity.

2. In a true-false examination the number of true statements should approximate the number of false statements, and the arrangement should be such that there is no regular sequence between true statements and false statements.

3. Since the pupil can give his responses very quickly, the examination should consist of not less than fifty statements. A true-false examination of one hundred statements can be given in the time usually devoted to an ordinary examination.

4. The examination should be mimeographed or printed so that each pupil will have a copy. He may give his answers in the margins of the sheets, or, if it is desired to use the same set of papers with another group of pupils, he may be given a sheet of paper on which there are numbered blanks. The pupils will then be asked to record in the blanks their answers to the corresponding exercises. A less desirable plan, which may be followed when it is not possible to secure mimeographed copies of the examination, is to read the statements to the pupils and have them record their answers in numbered blanks. The disadvantage of this plan is that the pupils do not have a satisfactory opportunity to study the statements. Also the class may give some indication of the answer if a statement appeals to them as being ridiculous.

5. The pupils should be given specific directions in regard to answering exercises about which they are uncertain. One writer

---

[6] For an example of a true-false examination, see Appendix p. 69.

has suggested that the pupils be instructed to guess concerning the truth or falsity of the statement. Another writer who has used this type of examination instructed the pupils as follows: "First, go through the list quickly and mark all that you know for certain, then go back and study out the harder ones. Do not guess; the chances are against you on guessing. Don't endanger your score by gambling on those questions about which you know nothing." This second procedure is probably the better.

The scoring of a true-false examination. Since only two responses are possible, it is obvious that a pupil may give a correct response as the result of chance. In order to take this possibility into account, a pupil's score on an examination of this type is the number of exercises answered correctly minus the number answered incorrectly. Exercises not attempted are not counted.

Directions for constructing a recognition examination.[7] In constructing this type of examination none of the proposed answers should be too obviously incorrect. An exercise can yield an indication of a pupil's achievement only when he is forced to use judgment in determining which of the proposed answers is suitable. For example, the illustrative exercise given would be practically worthless for testing purposes if all the names, except that of George Washington, were of persons living today or of persons having no connection with our national life. In applying this type of exercise to the field of arithmetic the proposed answers should include erroneous answers which pupils are inclined to give: if the exercise called for the quotient of two fractions, one of the proposed answers should be the product of the fractions and another their sum, and perhaps another should be the fraction obtained by taking the sum of the numerators as a new numerator and the sum of the denominators for a new denominator. When the correct answer is included in a group of such answers as these, the pupil who does not know how to find the quotient of such fractions will be unable to determine the correct answer except as a matter of chance. On the other hand, if all of the answers except the correct one were integers or were so large that they were obviously incorrect, a bright pupil who knew nothing about division of fractions would be able to select the correct answer. The correct answer should not always be found in the same position; sometimes it should be first, sometimes last, and sometimes in an intermediate position. As in the case of the true-false examination, a recognition examination should consist of a large number of exercises.

---

[7] For an example of a recognition examination see Appendix p. 75.

Examinations of this type should be mimeographed or printed and each pupil should have a copy. Definite instructions concerning methods of work should be given. It is probably best to instruct the pupil to work through the test rapidly, answering those exercises about which he is certain. He should then go back over the list and try the more difficult ones. Not fewer than four proposed answers should be included in each statement and the pupils may be instructed to guess if they do not know, since the chance of success by guessing is slight. The pupil's score on an examination of this type may be taken as the number of exercises done correctly.

A somewhat unusual but interesting type of recognition exercise is that described as a "matching contest." In this a pupil is given two lists of statements, the first numbered 1, 2, 3, 4, 5, etc., the second marked A, B, C, D, E, etc. In the second list, there is a statement which corresponds in meaning to a statement in the first list and the pupil is to pair these statements, marking by the number of the first list the letter of the corresponding statement of the second. For example, in the exercises given below: by the date marked (5), 1898, we place the letter B to indicate the event for which that date is significant. It is difficult to construct such examinations so that they will require reasoning on the part of the student. Their most important use is in the elementary school for rapid drill in certain phases of some subjects, such as definitions in geography and grammar, etc. The following exercises, selected from the Spokane United States History Test, illustrate the use of such an examination in linking a certain date or person with the corresponding event.

1. 1846
2. 1865
3. 1863
4. 1917
5. 1898
6. 1789
7. 1792
8. 1776
9. 1861
10. 1914

A. Lincoln's Emancipation Proclamation
B. Spanish-American War
C. Beginning of World War
D. Declaration of Independence
E. United States entered World War
F. Election of Washington as President
G. War with Mexico began
H. Invention of the cotton gin
I. Lee's surrender at Appomattox
J. Beginning of Civil War

1. Foch
2. Lincoln
3. Fulton
4. Dewey
5. Pershing
6. Bell
7. Edison
8. Jefferson
9. Lee
10. Franklin

A. Destroyed Spanish fleet in Manila Bay
B. Invented the telephone
C. Leading Confederate General
D. Wrote the Declaration of Independence
E. Invented the steamboat
F. Commanded allied armies in the World War
G. Was President during the Civil War
H. Commanded American Forces in the World War
I. Was Revolutionary patriot, author, and inventor
J. America's most famous inventor

Directions for constructing completion exercises.[8] A completion exercise should be constructed so that no suggestion will be given of the correct words to be written in the blanks. Furthermore, the facts to be supplied should be important. The best plan is to prepare a list of important statements and principles covering the portion of the subject over which the pupils are to be examined and then from these statements to strike out a certain significant word or phrase. In every case, if it is possible, the words omitted should be such that only one answer will be correct. Since little writing is required of the pupils they may be asked to fill in as many as one hundred blanks.

The scoring of completion exercises is not as highly objective as in the two types mentioned above. Pupils will tend to write a variety of words in the blanks. Different words may have almost the same meaning, and frequently the scorer will be compelled to determine whether the meaning of some word is sufficiently near that of the correct answer to justify giving the pupil credit for having answered the exercises correctly. However, by a careful selection of statements and of the omitted words, this subjectivity may be greatly minimized. For example, in the sentence, "The first Continental Congress was held in . . . .," only one possible word can be correct. In using completion exercises it is necessary to provide each pupil with a mimeographed or printed copy of the examination. The pupil's score is the number of blanks filled in correctly.

Advantages of the "new examination." Examinations consisting of exercises of the types described above have certain obvious advantages. There will be a large saving of time for both teacher and pupil. The pupil is called upon to do little or no writing in giving his answers and therefore is able to respond to a large number of exercises. The teacher in scoring will have little or no occasion to use judgment as he will need only to note the brief responses given by the pupils. Thus the labor of scoring will be greatly reduced and, more important, the scoring will be much more highly objective than that in the marking of examination papers of the usual type. The saving of time in the giving and scoring of the "new examination" will more than offset any additional time that may be expended in its construction. Another advantage is that the new examination can be made more comprehensive. Examinations as a rule consist of ten questions. Some are limited to a smaller number. Consequently the scope of examinations of the traditional type is necessarily narrow. "New examinations" of the true-false type should consist of not less than fifty exercises and may have as many as one hundred. Other types of the "new examination" should be of a corresponding length. Hence a "new examination" will usually be more comprehensive than a traditional examination.

---

[8] For an example of a completion examination see Appendix p. 73.

Limitations of the "new examination." There are certain limitations of the new examination which should be noted along with its advantages. It can not be used in mathematics except to a limited extent. It can not be used at all in English Composition. In other subjects there are many phases of achievement which are not measured directly by examinations made up of exercises of the types described above. Hence, altho the "new examination" is more comprehensive with reference to information, and does measure certain types of achievements, it is likely that pupils would miss much valuable experience and training if they were not at times asked to "compare," "explain," "discuss," "define," or "tell why." They should also be asked to summarize material presented on a topic or to apply certain principles. The following questions taken from Hahn's Scale for Measuring Ability of Children in History appear to require mental processes distinctly different from those called for by the "new examination."

1. "State points of similarity between the position of the United States in 1812 and their position in 1912."

2. "Arrange the following events in order of cause and effect: Force Bill, Carpet Baggers, Fifteenth Amendment, Negro Rule in Some of the Southern States, Ku Klux Klan."

3. "Name the presidents of the United States since 1892."

An intelligent attitude toward the "new examination." The simple administration of the new examination and other attractive features should not blind one to the limitations just mentioned. As indicated in Chapter II written examinations do more than merely secure measures of achievement. If they consist of the right kind of exercises they afford significant educational opportunities. The educational opportunities of the "new examination" are necessarily restricted, and it would be unfortunate if it entirely replaced examinations of the traditional type. The new examination, however, has a place. It may be used occasionally in most school subjects. It is useful when a teacher wishes to test the acquaintance of a class with a wide range of facts. It has little diagnostic value and examinations of the traditional type should be used when information is desired concerning the weaknesses of different members of a class. For this reason the "new examination" is more appropriate for use at the end of a term than for tests during the term which have as their purpose both measurement and diagnosis.

CHAPTER VII

RULES FOR THE PREPARATION AND ADMINISTRATION OF WRITTEN EXAMINATIONS.

Below, a group of suggested rules governing the preparation and administration of written examinations are given. These represent the opinion of the writers which is based upon a careful study of the problems involved, as well as upon several years of experience in the measurement of school achievement.

1. Final examinations should be required. In school subjects such as shop work, in which the performances secured from pupils are highly objective, the waiving of this requirement may be justified. When final examinations are given no student should be excused from them because of high daily grades, deportment or attendance. (See p. 16)

2. The content of final examinations should agree as closely as possible with recognized educational objectives. In fields where minimum essentials have been determined they should be used as a basis in formulating questions. (See p. 55)

3. The questions should be definite and stated so that all pupils will interpret them alike. Questions relating to items of minor importance should occupy a minor place in examinations. Questions relating to points which have not received attention in the course should be omitted.[1] (See also rule 11.)

4. When the necessary equipment is available the questions should be mimeographed or typewritten so that each student will have a copy on his desk. In case they are written on the board the teacher should make certain that all pupils are able to read them correctly. It is well in either case to read the questions aloud to the class.

5. The examination should be sufficiently difficult so that few pupils will make perfect scores. (This rule should not apply

---

[1] Frequently in an examination the difficulty of the question is due to the lack of emphasis placed upon it throughout the term because it deals with a relatively unimportant topic. Other topics, difficult in themselves, but emphasized because of their importance, furnish the easier questions.

when the purpose of the examination is to determine which students have attained a given standing which includes perfection of performance.) (See p. 12)

6. Usually the examination should be long enough so that every member of the class will be kept busy for the entire period. It is better to make the examination too long than to have it too short. In this way it becomes possible also to take into account the student's rate of work in determining his grade. Appropriate adjustments can be made in interpreting "scores" into school "marks." (See p. 22 and 13)

7. In questions asking for a discussion or explanation indicate the completeness of the discussion or the degree of elaborateness expected in the answer. (See p. 24)

8. Time may be economized for both students and teacher by using some form of the "new examination." This type of measuring instrument, however, possesses certain limitations which should be kept in mind. The exclusive use of it would be unwise. (See p. 61)

9. Unless the students have a definite understanding of the methods of work which are to be followed, the teacher should give them explicit directions concerning such matters as the order in which the questions are to be answered, the desired arrangement of the work, and any other items in which there is an opportunity for pupils to adopt different procedures. (In the case of most standardized educational tests, the directions to students are very detailed and explicit.) (See p. 24)

10. Approximately ninety minutes should be allowed for a final examination in most high-school subjects. The time which teachers should devote to the preparation of the questions will depend upon their experience and upon their practise during the semester. It is recommended that a teacher make a record throughout the term of questions which in his judgment are suitable for a final examination. From two to three hours should be sufficient for marking a set of twenty-five examination papers. If a teacher finds that a longer time is required, he should endeavor to modify his procedure so that this work can be done more quickly. (See p. 19)

11. Altho any weighting of questions by a teacher will be subjective, it is probably desirable to weight the questions, particularly in cases of extreme differences in value. The weighting,

however, should be upon the basis of social importance rather than upon mere difficulty. (See p. 22)

12. It is advisable for the teacher to write out at least in an abbreviated form the answers to examination questions before he begins marking the papers. In mathematics and other subjects where a definite answer is required and only one can be accepted as correct, the need for this rule is not as great as in such subjects as geography, history, literature and certain phases of science. However, a list of correct answers will usually mean a saving of time. (See p. 23)

13. Except in courses in English a pupil's grade should not be intentionally lowered for errors in spelling or for poor handwriting. As a rule the grade should not be lowered because of poor English, unless the quality of the English is evidence of unsatisfactory reasoning. Rules covering these points, as well as others concerning which teachers might differ, should be formulated by the principal in conference with his teachers or at least a committee of them. The rules thus formulated should be carefully followed by all of the teachers. (See p. 20)

14. In marking the papers more accurate results will in general be secured if the answers to one question are marked on all the papers before those for another question are taken up. (See p. 23) When it is desired to mark all of the questions on one paper before taking up another, the "sorting method" should be used. According to this procedure the papers as they are read are sorted into piles, the best ones being placed in the first pile, the next best in second pile, etc. Five distributions will, in most cases, prove sufficient. After all the papers have been distributed they should be reread, one pile at a time, and compared with each other. If these papers do not possess approximately the same value, changes in the sorting may be made. Grades may then be assigned to the papers in the different piles.

15. The distinction between "scores" and "grades" should be kept in mind. (See p. 11) The papers should be marked first in terms of scores. In doing this an appropriate number of points should be determined for each question. It is not necessary that the total of these points be 100. (See p. 52)

16. The point scores assigned to the examination papers should be translated into school marks. In doing this the use of a standard distribution will be found helpful, and will operate also to decrease the magnitude of the constant errors. (See p. 50)

APPENDIX

(Questionnaire Sent to High-School Teachers)

QUESTIONNAIRE RELATING TO THE USE OF WRITTEN EXAMINATIONS IN HIGH SCHOOLS

Teacher..High School.......................................City..........................

The major subject which I am teaching is...

The following questions are to be answered only with reference to the major subject you are teaching.

1. Approximately how much time do you use in preparing questions for a final examination which the students are allowed a total of 90 minutes to answer?.. ....Min.

("Final examination" as used in this questionnaire means an examination which is given at the end of a semester and which is based on the work of the entire semester.)

2. In preparing a set of examination questions do you usually attempt to arrange these questions in order of ascending difficulty?............ Yes No*

3. Do you prepare in written form carefully worded directions to the students regarding the procedure they are to follow in answering the questions? (These directions might include such points as, order in which questions are to be answered, length of answers, arrangement of work, etc.) Yes No

4. Which of the following methods do you use in presenting questions to the student?...

(a) Writing the questions on the board.. Yes No

(b) Furnishing the pupils with a mimeographed, carbon or printed copy of the questions.. Yes No

(c) Dictating the questions to the pupils.................................... Yes No

5. Is it your custom to make the examination long enough so that practically none of the students will answer all of the questions in the time allowed?.. Yes No

---

*Underline the answer (Yes-No) which you desire to make.

6. Is it your custom to make the examination short enough so that practically all of the students will answer all of the questions in the time allowed?............ Yes No

7. If you have given an affirmative answer to Question 6, do you note the time each student spends in writing his answers?............ Yes No

8. What proportion of the final mark for the semester is based upon the final written examination grade?............ ....%

9. In assigning grades to examination papers do you attempt to have their distribution conform to any standard form such as the normal distribution?............ Yes No

10. Do you usually grade all the answers on one paper before taking up those of another paper?............ Yes No

11. Do you usually grade the answers to one question on all of the papers before taking up the answers to a second question?............ Yes No

12. Instead of marking the answers to each question separately do you attempt to estimate the value of the paper as a whole?............ Yes No

13. Before starting to grade a set of examination papers do you write out the answers which you consider correct?............ Yes No

14. When you consider the questions of an examination to be unequal in difficulty is it your practise to give more credit for a correct answer to a difficult question than for a correct answer to an easy question? Yes No

15. Approximately how much time do you use in marking the papers of a final examination which the students are allowed a total of 90 minutes for answering? Estimate as accurately as possible. Base this answer on a class of 25 students............ ....Min.

16. In marking examination papers do you intentionally lower a student's mark in the case of

(a) poor writing............ Yes No
(b) poor spelling............ Yes No
(c) poor English............ Yes No

17. In the case of questions which are essentially mathematical in character do you give credit to the student for using the correct principle even though the final answer be wrong?............ Yes No

NOTE: It is desired that only teachers of mathematics, physics, and chemistry answer No. 17.

NOTE TO TEACHER: When you have answered the above questions please return this questionnaire to your principal.

(Questionnaire Sent to High-School Principals)

Principal....................................High...City............................

1. Do you require your teachers to give final examinations at the end of each semester?.. Yes No*

("Final examination" as used in this questionnaire means an examination which is given at the end of a semester and which is based on the work of the entire semester)

If a *negative* answer is given to the first question, no answer is expected for the remaining questions.

2. a. Is it the practise in your school to exempt certain students from final examinations?.. Yes No

b. If so, what are the conditions (requirements) upon which you base exemption?

1. Deportment.. Yes No
2. Scholarship.. Yes No

Other requirements.

3. How many minutes do you allow for final written examinations?.... ....Min.

4. Because final examinations have proven unreliable some educators urge that students be given more than one comprehensive examination in each subject, and that these examinations be given on different days. Do you require more than one such final written examination in each subject?.. Yes No

5. a. In marking examination papers is it the practise in your school for the teachers to subtract from a pupil's grade for

1. poor writing.. Yes No
2. poor spelling.. Yes No
3. poor English.. Yes No

b. Are your teachers accustomed to giving more credit for correct answers to difficult questions than for correct answers to easy questions?.. Yes No

c. Is it the practise of your teachers when computing a semester mark to add or deduct credit in proportion to the time used by a pupil in answering the final written examination questions? Yes No

6. a. Have you advised your teachers as to what proportion of the final mark for the semester should be based on the final written examination?.. Yes No

b. Have you made a definite requirement in this respect?........... Yes No

c. If so, what proportion of the final mark for the semester do you require to be based upon the final written examination mark?.. ....%

7. Additional information pertaining to this topic is called for in a a second questionnaire which is to be filled out by high school teachers. Would you be willing to distribute this questionnaire among your teachers? If so, will you kindly indicate the number of teachers in your high school?.. ....

---

*Underline the answer (Yes-No) which you desire to make.

EXAMPLES OF "NEW EXAMINATIONS"

(The following "new examinations" are given for purposes of illustration. They may include several exercises which will prove unsuitable when given to pupils.)

TRUE-FALSE EXAMINATION IN PHYSIOLOGY

Prepared by

Bureau of Educational Research

University of Illinois

Name..Boy or Girl....................................

Age last birthday........................Next birthday will be..19........

Grade........................Date...............................City...............................State..........................

School...Teacher...

Below you will find a number of statements. Some of these statements are true, others are not true. Read each statement carefully, then if it is true mark a plus (+) in the column to the right of the sentence. If the statement is not true mark a minus (−) in the column to the right.

EXAMPLES

Answers to be written here

Read the statement below very carefully.

1. **Fats will form a lasting mixture with water.**

This is not a true statement so you will place a minus (−) sign in the column. Now read the second sentence.

2. **The layer of fat just beneath the skin is more than one-tenth of an inch thick.**

This is a true statement so you will mark a plus (+) sign in the column. Now read the third sentence.

3. **The union of oxygen with any substance produces heat.**

This is a true statement so you will mark a plus (+) sign in the column. Now read the fourth sentence.

4. **Nitrogen constitutes only one-fifth of the volume of the air.**

This is a false statement so you will mark a minus (−) sign in the column..

PHYSIOLOGY

Answers to be written here

1. The kidneys vary 12 inches to 16 inches in length.
2. The external poisoning of the skin by poison ivy or sumac never results seriously.
3. A person having a good mind must necessarily have a large brain.
4. Color blindness is more prominent in men than women.
5. Plenty of fluids should be drunk at the time of eating solid food.
6. Bones are composed of animal and mineral matter.
7. The nails are hardened outer skin or epidermis.
8. The use of alcohol increases the tendency to commit crime.
9. A full grown person contains about six quarts of blood.
10. The brain is almost perfectly spherical in shape.
11. The kidneys are almost perfectly round.
12. All animals are made up of cells.
13. Substances, like glass, which permit rays of light to pass through them readily are said to be opaque.
14. The sense organs of smell are located in the lining of the cavity of the nose.
15. The skin is composed of two layers of tissue.
16. The end organs for taste occur in the mucous membrane of the tongue.
17. To extinguish the burning clothing of a person, it is necessary to wrap him in something to exclude the air.
18. All of the interior of the spinal cord is filled with gray matter containing nerve cells.
19. The great difference in the complexion of persons is due largely to the pigment lying in the epidermis.
20. Cancer is caused by germs growing in the tissue.

Answers to be written here

21. Diphtheria can be controlled by the use of Diphtheria antitoxin.

22. The brain is separated into two parts or hemispheres by a great longitudinal fissure.

23. When oxygen is separated from other substances the process is called oxidation.

24. Infectious diseases are due to changed methods of work and growth on the part of cells in certain regions of the body.

25. The use of alcoholic beverages builds up the body and makes the muscles stronger.

26. The great majority of grown people have been infected with tuberculosis germs.

27. The sense organs are the terminations of the sensory nerves serving to carry impressions to the spinal cord or brain.

28. Farsightedness is often caused by a blow on the eye.

29. An antiseptic is a substance which merely restrains the germs from growing.

30. The brain is in communication with the rest of the body by means of nerves.

31. The cerebrum is the path of communication between the nerves supplying the arms, trunk, legs, and brain.

32. The chief function of muscles is to hold up the body.

33. All milk contains bacteria.

34. The alcohol used in drinks is produced by the growth of yeast in a liquid containing sugar.

35. Our blood contains white corpuscles which destroy disease germs.

36. More people die daily from diphtheria than from tuberculosis.

37. The use of tobacco increases the strength of the muscles.

38. The use of tobacco makes the nerve cells function more keenly.

39. The chewing of dry bread aids the digestion as much as the use of gum.

	Answers to be written here
40. Air is composed chiefly of two gases, oxygen and nitrogen.	
41. The first step in treating a person who has been poisoned is to give an emetic.	
42. Light is produced by waves of a substance called ether.	
43. Non-infectious diseases are caused by small plants or animals called parasites feeding upon the human body.	
44. Alcoholic beverages have great value in curing disease.	
45. A drink of alcoholic beverage in the winter time causes a man's body to become warm.	
46. Each portion of the brain has its own definite work to perform.	
47. Fainting is caused by an over-sufficient supply of blood being sent to the brain.	
48. The spinal cord may act independently of the brain and produce many of the muscular movements necessary in routine work.	
49. The germs of typhoid fever usually gain access to the body by being breathed in with air.	
50. Narcotics are substances which cause any organs of the body to act more vigorously than is their custom.	

Directions to teachers: After the four examples have been studied by the pupils, read the following directions to them: "On the next page you will find a number of statements similar to the ones you have just read. You are to place a plus sign or a minus sign in the column to the right of each statement just as has been done on the first page. Mark all of the statements that you are sure you can answer correctly. If you find a statement that you are not sure you can answer correctly, study it carefully and then mark the answer you think will be correct. If you find a statement you know nothing about, make no attempt to mark it, as guessing counts heavily against you. You will have 25 minutes for the test. I shall expect you to stop promptly and turn your folders face down on the desk when I tell you to do so. Ready-Go."

In computing the score of each pupil on a test subtract the total number of wrong answers from the total number of right answers. Such scores are called "point-scores." In interpreting them it is advisable to form a distribution which will show how many pupils received each score. From the distribution it is possible to work out a basis for translating the point scores into the usual kind of school marks.

NOTE: The "Directions to teachers" given above would not appear on the usual printed examinations. They are placed here for the convenience of teachers.

COMPLETION EXAMINATION IN AMERICAN GOVERNMENT

Prepared by

Bureau of Educational Research

University of Illinois

Name..Boy or Girl....................................

Age last birthday.......................Next birthday will be...19........

Grade.......................Date...............................City..........................State............................

School...Teacher..

Below you will find a number of statements. In each statement one or more important words have been omitted. Each blank in the sentence shows where a word has been left out. Read each statement carefully, then write in the blank the word which completes the meaning of the statement. You will be allowed 15 minutes for the test.

1. The primary purpose for which government exists is the.......................of our lives and property.
2. Citizenship may be acquired by.......................in this country or by a process of ....for natives of other lands.
3. Our national government derives its authority from the.......................of the United States through our national...............................
4. The legislative power granted to the national government is vested in a Congress of ....houses, the smaller of which is called the..............................and the larger the..........................
5. The execution of the laws made by.......................is intrusted to the.......................... of the United States.
6. All judges connected with the national courts are appointed for life with the consent of...
7. Most of the candidates for office which are filled by popular vote are nominated directly in......................................
8. The Fifteenth Amendment of the United States Constitution prevents the states from denying to citizens the right to vote on account of..., ..., or previous condition of..............................

9. Practically all of our law-making bodies are made up of................................chosen for short terms from........................into which the states, counties and cities are divided.

10. The first permanent English settlements in America were made in what is now the state of...

11. In a county, the records of the county board and other official papers are preserved by the county...

12. All cities are public corporations created under...........................municipal laws.

13. Every incorporated city obtains from the........................government a.....................under which it may elect its officials and conduct its business.

14. Civil service employees may be removed from service only for................................

15. The power of impeaching a state officer is given to the...

16. The...............................is by far the most prominent and powerful executive official in the state. Very........................state officers are appointed by him or are responsible to him.

17. All important officials connected with the executive or judicial service of the United States may be removed by...............................through the lower house of Congress and by...............................in the senate.

18. Far more property is destroyed by...........................than by all other agencies.

19. There is no task of state and local government which outranks in importance that of providing an......................................education at public expense.

20. All rivers and canals within a single state are controlled by the...............................in which they are located.

21. Most of the revenue for state and local governments is secured by a.........................on...........................

22. A state.......................is the fundamental law which the people of the state have arranged for their government and protection.

23. A state constitution can be changed by means of an...

24. The three-fifths compromise provided that five.......................should be counted as equal to three.......................when reckoning the...............................for either direct taxation or representation.

RECOGNITION EXAMINATION IN ALGEBRA

Prepared by
Bureau of Educational Research
University of Illinois

Name..Boy or Girl...

Age last birthday........................Next birthday will be...19........

Grade........................Date........................City...........................State.......................................

School...Teacher..

Below you will find a number of statements. In each statement a word or number has been omitted. At the close of the statement several words or numbers have been given. One of these is the correct answer. Select the word or number which you think is correct and draw a line under it. Most, if not all, of the examples can be solved by mental calculation. If any figuring is necessary, work on the margin of the page. You will be allowed 17 minutes for the test.

1. Numbers that are represented by letters are called............................numbers.
   substituted—literal
2. When two or more letters are multiplied together each is called a...............................
   of the product. factor—coefficient
3. If a man rides a certain distance in 10 hours, in h hours he rides...............................
   $10h$; $\frac{h}{10}$; $\frac{10}{h}$
4. The statement $2x + 5 = 29$ is called an........................ identity—equation.
5. If 16 is subtracted from three times a certain number the result is 110. The number
   is........................ $36\frac{2}{3}$; $31\frac{1}{3}$; 42
6. A number which is a factor of two or more numbers is called a........................factor.
   common—equal
7. If there are two equal factors of a number, either is called the........................of the
   number. square root—common factor
8. To multiply algebraic fractions take the..............................of the numerators for a
   new numerator and the product of the denominators for a new denominator.
   sum—product

9. A fraction whose numerator or denominator (or both) contains fractions is called a ....fraction. multiple—complex

10. A.......................is a statement of a fact which is to be proved. theorem—axiom

11. The name given the + sign is.......................... negative—positive

12. To find the sum of two numbers whose signs are opposite, take their....................... regarding each as positive, and prefix the sign of the larger number to the answer. sum—difference—product

13. Whenever a number occurs without a sign, the..........................sign is to be understood. ×; +; −

14. The number denoting the power of a term is called the.. prefix—exponent

15. If $a = 2$, $b = -3$, and $c = -5$ then $\frac{2abc}{a} =$ .... -6; -2; 30

16. In adding like terms add the coefficients for the new coefficient and....................... it by the common factor. multiply—divide

17. An expression which contains more than one term is called a.................................. monomial—polynomial

18. If the length of a rectangle = 4 feet more than twice the width, the perimeter = 56 feet. The length =.......................feet. 8—12—16

19. Any term may be transposed from one side of an equation to the other, provided its ....is changed. sign—value

20. Any equation which contains no higher power of the unknown letter than the first is called a..............................equation. radical—simple

21. The exponent of the product of two powers of the same number is equal to the ....of the exponents of the factors. product—sum

22. To raise the product of two numbers to any power, raise the numbers separately to that power and take their....................... product—sum

23. The square of any two numbers is equal to the square of the first number............... twice the product of the two plus the square of the second number. plus—minus

24. A 20 foot ladder rests against a building, the bottom of the ladder being 12 feet from the cellar wall. The top is...............feet from the ground. 8-16

25. In division, the sign of the quotient is.......................whenever the dividend and divisor have like signs. −; +

26. In finding the quotient of two powers of the same number the exponent of the quotient is equal to the exponent of the dividend....................by that of the divisor.
increased—diminished

27. $(3x^2-2x-1)\div(x-1)=$........................ $3x+1; 3x-1$

28. A factor which has no factor except itself and unity is called a.......................factor.
prime—multiple

29. The product of all the common prime factors of two or more numbers or expressions is called their............................common factor. highest—lowest

30. If one number is exactly divisible by another, the first is called a....................of the second. divisor—multiple

31. In algebraic fractions the dividend is called the...................................
denominator—numerator

0

LB 1028 A1I45 no.2-17
Illinois. University. Bureau of Educational Research
Bulletin

PLEASE DO NOT REMOVE
CARDS OR SLIPS FROM THIS POCKET

UNIVERSITY OF TORONTO LIBRARY